42 No. 1 (New Series No. 192) January 198

Cuk-H02964-41-P00D615

ANALYSIS

Edited by
CHRISTOPHER KIRWAN

11

CONTENTS

Metaphysical realism — J. J. C. SMART
Vague objects — HAROLD W. NOONAN
Sentence modifiers and semantic theories: a note on a conjecture — MARTIN DAVIES
Translation and belief — ARNOLD CUSMARIU
The concept of experience and Strawson's transcendental deduction — KIM DAVIES
Mackie's defence of induction — P. J. R. MILLICAN
Chisholm's objection to phenomenalism — MICHAEL J. MALONEY
The paradox of the 1,001 cats — E. J. LOWE
Reply to Lowe — P. T. GEACH
Reply to Geach — E. J. LOWE
Reply to Lowe's reply — P. T. GEACH
Canonical counterpart theory — GRAEME FORBES
What won't escape sorites arguments — PATRICK GRIM
Personal identity and personal survival — ANDREW BRENNAN
Can we travel faster than light? — CHRISTOPHER RAY
A note on the synonymy principle of property identity 52 — MICHAEL TYE
Utilitarianism and rationality — EARL CONEE
The argument for anomalous monism — TED HONDERICH

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003-2638

The Philosophical Quarterly

Special Numbers:

The Editors intend to devote one issue in 1982 to articles on **Scientific Realism** and one issue in 1983 to articles on **Medical Ethics.** Contributions should be submitted by **31 December of the preceding year,** to the Executive Editor, The Philosophical Quarterly, The University, St Andrews, Scotland.

Vol. 32 No. 126 *January 1982*

Intention, Responsibility and Double Effect	*Antony Duff*
Knowledge and Understanding	*John Campbell*
The Value of Autonomy	*Robert Young*
Towards an Eliminative Reduction of Possible Worlds	*Andrew Roper*
DISCUSSIONS	
Bad Samaritans Cause Harm	*John Harris*
Lowe on Intentionality	*Eric Stiffler*
Gewirth on Absolute Rights	*Jerrold Levinson*
CRITICAL STUDY	
Jonathan Lear, *Aristotle and Logical Theory*	*John Corcoran and Michael Scanlan*
BOOK REVIEWS	BOOKS RECEIVED

Annual Subscriptions:
Individuals £8.00 ($20.00) Institutions £16.00 ($40.00)
for orders placed directly with the Journals Manager, Scottish Academic Press Ltd., 33 Montgomery Street, EDINBURGH EH7 5JX, Scotland.

P2615

RATIO

edited by
MARTIN HOLLIS

Volume XXIII *December 1981* *No. 2*

CONTENTS

GOTTFRIED SEEBASS	*Mediation Theory and the Problem of Psychological Discourse on 'Inner' Events: Part I*
FRANCIS W. DAUER	*Incorrigibility*
CLARK ZUMBACH	*A Note on the Thesis Argument of the Third Antinomy*
KEITH GRAHAM	*How do Illocutionary Descriptions Explain?*
ROY A. SORENSEN	*Disagreement amongst Ideal Thinkers*
ALAN N. SUSSMAN	*Counter-Factuals and Non-Trivial* De Re *Modalities*
RICHARD SHUSTERMAN	*Evaluative Reasoning in Criticism*
JONATHAN HARRISON	*Gulliver's Adventures in Fairyland*

Books Received

Subscriptions
Institutions: £18.50 (UK) £22.75 (overseas) $49.50 (USA) $57.00 (Canada)
Individuals: £12.50 (UK) £15.25 (overseas) $32.50 (USA) $37.25 (Canada)

Orders to: Journals Department, Basil Blackwell Publisher Ltd.,
108 Cowley Road, Oxford OX4 1JF, England.

The German edition of *Ratio* is published by FELIX MEINER VERLAG,
Richardstrasse 47. 2000 Hamburg 76, West Germany.

Published by BASIL BLACKWELL, Oxford

ANALYSIS

Edited by Christopher Kirwan

105
Ar 13

Volume 42

1982

ISSN 0003-2638

BASIL BLACKWELL · OXFORD

INDEX

		PAGE
ANALYSIS Problem No. 19		115
ANGENE, LYLE E.	Five o'clock here	78
BALDWIN, THOMAS	Sets whose members might not exist	133
BARNES, GERALD W.	Mince pie reasoning	163
BRENNAN, ANDREW	Personal identity and personal survival	44
CARTER, W. R.	Omnipotence and sin	102
CONEE, EARL	Utilitarianism and rationality	55
CUSMARIU, ARNOLD	Translation and belief	12
DALE, A. J.	Logical equivalents and logical form	190
DAVIES, KIM	The concept of experience and Strawson's transcendental deduction	16
DAVIES, KIM 105 13	Intentionality: spontaneous ascription and deep intuition	169
DAVIES, MARTIN	Sentence modifiers and semantic theories: a note on a conjecture	7
EHRING, DOUGLAS	Manipulability theory and event types	149
ELLIOT, ROBERT	Going nowhere fast?	213
FOGELIN, ROBERT J.	Wittgenstein's operator *N*	124
FORBES, GRAEME	Canonical counterpart theory	33
GEACH, P. T.	Reply to Lowe	30
GEACH, P. T.	Reply to Lowe's reply	32
GEACH, P. T.	More on Wittgenstein's operator *N*	127
GODWIN, WILLIAM	Analysis and the identity of indiscernibles	80
GOTTLIEB, A. J.	Imperceptible parts	111
GRAYLING, A. C.	Internal structure and essence	139
GRIM, PATRICK	What won't escape sorites arguments	38
GRIM, PATRICK P2615	Against a deontic argument for God's existence	171
HAZEN, ALLEN	On a possible misinterpretation of Kripke's semantics for intuitionistic logic	128
HONDERICH, TED	The argument for anomalous monism	59
HORNSBY, JENNIFER	Reply to Lowe on actions	152
HORNSBY, JENNIFER	A thesis refutable by a sentence verifiable by its use	177

HUMBERSTONE, LLOYD	(see JACKSON, FRANK)	
HUNTER, GRAEME	(see SEAGER, WILLIAM)	
HURKA, T. M.	Average utilitarianisms	65
HURKA, T. M.	More average utilitarianisms	115
JACKSON, FRANK *and* HUMBERSTONE, LLOYD	On a challenge by Anderson and Belnap	179
KEATING, B. F.	Rowe, self-existence, and the cosmological argument	99
KING, JEFFERY	'Five o'clock on the sun': a reply to J. L. Mackie	77
LOWE, E. J.	The paradox of the 1,001 cats	27
LOWE, E. J.	Reply to Geach	31
LOWE, E. J.	Intentionality and intuition: a reply to Davies	85
LOWE, E. J.	On being a cat	174
MALONEY, MICHAEL J.	Chisholm's objection to phenomenalism	25
MELLOR, D. H.	Counting corners correctly	96
MILLER, BARRY	Negative existential propositions	181
MILLICAN, P. J. R.	Mackie's defence of induction	19
MILLICAN, P. J. R.	On the transitivity of epistemic preferability	88
MONTMARQUET, JAMES A.	Causal deviancy and multiple intentions	106
MORE, M. J.	Rigidity and identity across possible worlds	83
NATHAN, N. M. L.	Conscious belief	90
NOONAN, HAROLD W.	Vague objects	3
NOONAN, HAROLD W.	Williams on 'The Self and the Future'	158
OVER, D. E.	Predicative and constructive knowledge	140
PRIOR, ELIZABETH W.	The dispositional/categorical distinction	93
RASMUSSEN, STIG ALSTRUP	Ruben on Lewis and causal sufficiency	207
RAY, CHRISTOPHER	Can we travel faster than light?	50
SEAGER, WILLIAM *and* HUNTER, GRAEME	Reply to Forbes	224
SIMONS, PETER M.	Token resistance	195
SLOTE, MICHAEL	Is virtue possible?	70
SMART, J. J. C.	Metaphysical realism	1

SMITH, PETER	In defence of Lewy	119
SMITH, PETER	Bad news for anomalous monism?	220
SORENSEN, ROY A.	Knowing, believing, and guessing	212
STAHL, DONALD E.	Goldman on what justifies belief	146
STEINER, HILLEL	Vanishing powers: a reply to Miller and Wilson	97
TAYLOR, DENISE MEYERSON	Actions, reasons and causal explanation	216
TYE, MICHAEL	A note on the synonymy principle of property identity	52
WARD, ANDREW	Materialism and memory	153
WELBOURNE, MICHAEL	My body and I – a reply to Fahrnkopf	86
WILLIAMS, C. J. F.	Reply to Miller	189
WILLIAMSON, T.	Intuitionism disproved?	203

ANALYSIS 42:1 JANUARY 1982

METAPHYSICAL REALISM

By J. J. C. SMART

HILARY PUTNAM in his 'Models and Reality' has distinguished 'metaphysical realism' from 'empirical realism' (*Journal of Symbolic Logic* 45, 1980, pp. 464–82). An empirical realist admits that reference is 'trans-theoretic' in the sense that a later theory, incompatible with our present ones, may come to be preferred as a result of further observation and theoretical enquiry. A metaphysical realist, according to Putnam, goes much further and holds that it is coherent to suppose that an *ideal* scientific theory, which satisfied all possible operational and theoretical constraints, could nevertheless be false (ibid., p. 473). It seems that Putnam thinks of truth and reference as in a sense theory-relative, so that if an actual theory T_i is regarded as possibly false, this is only because it could be seen as false from the point of view of a succeeding theory T_{i+1}.

Putnam's arguments against metaphysical realism are complex and subtle (involving discussion of the Löwenheim-Skolem theorem) and it is not my present purpose to examine them in detail. Instead I wish to cast doubt on them by questioning the truth of their conclusion. I wish to give a particular example of how an ideal theory T_I could consistently be supposed to be false. Before doing so a few words are needed about how T_I is specified.

Putnam distinguishes a set of all possible 'operational constraints' *OP* from a set of 'theoretical constraints'. The set *OP* is a valuation that assigns the correct truth value to each n-place observation term over each n-tuple of observable entities, where 'observable' is glossed by Putnam as 'observable with the human sensorium' (ibid., p. 472). Thus any actual observational test of a theory falls short of satisfaction of *OP*, since *OP* corresponds to infinitely many observations carried out all over the physically accessible universe. The theoretical constraints consist of such principles as simplicity and Occam's razor. Then the ideal theory T_I is characterized as satisfying both all the operational and all the theoretical constraints. Putnam glosses T_I as 'epistemically ideal for humans'.

Now for my example. Let us suppose that the ideal theory T_I does not differ from the solid core of present-day science in postulating that all physical entities belong to a four-dimensional space-time manifold. Now consider another theory, T_I^*, that is like T_I in what it says about the contents of our space-time manifold, but that asserts that our space-time manifold is a cross-section of a five-dimensional manifold which contains as another cross-section another space-time manifold, also, like ours, containing galaxies, stars, planets, and so on. Let us suppose that according to T_I^* we can never know about this other space-time sub-universe because this sub-universe and our own one are precluded by the laws of nature from causally interacting with one another.

The second space-time sub-universe envisaged by T_I^* may be thought to be worryingly like David Lewis's realistically conceived 'possible worlds' (see for example David Lewis *Counterfactuals*, Blackwell 1973). Like these, the other sub-universe is not in any physical interaction with ours. Now there are at least two reasons for not believing in Lewis's possible worlds. One is simply Occam's razor. (It might be said that Occam's razor is not applicable because the possible worlds are needed for the interpretation of ordinary language – contrary to fact conditionals, modal language, and so on. But if the web of belief can be developed without this part of language we can say at the most that a lot of scientifically dispensable ordinary language implicitly asserts the existence of possible worlds other than the actual world. I myself would think that either modal language – counterfactuals etc. – must be rejected as superstitious or else some account must be possible without bringing in possible worlds. Occam's razor is too persuasive here!) The other (possible) objection to Lewis's possible worlds semantics comes from difficulties over the counterpart relation between objects in different worlds. Now it seems that though Occam's razor can be applied to T_I^* in favour of T_I, the other sort of difficulty does not arise. We are not concerned with modal semantics and the question of whether a galaxy in the other T_I^* world is or is not a counterpart of a galaxy in ours simply does not arise.

We would never have reason to believe T_I^* as opposed to T_I (even assuming, per impossibile, that we could know all the operational constraints) because T_I^* violates one of the theoretical constraints, namely that of simplicity. Nevertheless it seems perfectly intelligible and consistent to suppose that after all T_I^* might be true. The universe might be less simple than we hope it is, so that our theoretical constraint of simplicity leads us astray. Certainly according to our present standards of rationality a hypothesis such as T_I^* seems (where it differs from T_I) an idle and gratuitous one, that no rational person would accept. But the universe might trick even an infinite rational being who knew all of *OP* and correctly applied the most rational of methodological canons. All this would lead him to T_I, and T_I^* might be true.

There is a question of whether if T_I^* were true *OP* should contain possible observations in the other space-time manifolds, so collapsing T_I^* into T_I. But if T_I is the rationally ideal theory, it could not be held by a group of people that straddled both space-time manifolds. Surely rationality belongs to a single person or intercommunicating group of persons. Alternatively we could fix up the story by supposing that in the other region there are no sentient beings (or even macroscopic objects – everything might be a plasma). Of course it is unlikely that another universe broadly similar to ours would not contain sentience, but as I am supposing a rejection of present theoretical constraints I may as well suppose the unlikely. So I do not *think* that an extension of *OP* into the other manifold arises as a problem for me.

Our present science contains enough for us to have conferred meaning on talk of a four-dimensional space-time manifold, in which we ourselves figure as 'space-time worms'. Current physical theory also contains a great deal of mathematics, and I agree with W. V. Quine that this is an integral part of physics and is to be believed because we believe physics. But having got our mathematics, with its Platonist ontology, we can make use of this mathematics and its ontology to describe the hypothetical five-dimensional universe that is envisaged by T_I^*.

I concluded (a) that it would never be rational to believe T_I^* but that (b) it might nevertheless be the case that T_I^* is true. In short metaphysical realism is true.

The Australian National University ©J. J. C. SMART 1981

VAGUE OBJECTS

By HAROLD W. NOONAN

GARETH EVANS' paper 'Can there be vague objects?' (ANALYSIS 38.4, October 1978) refutes the idea that an identity statement 'a = b' might be indeterminate in truth-value because one or both of a and b might be an object about which it is a *fact* that it has fuzzy boundaries. The only other way 'a = b' might be indeterminate in truth-value is by either 'a' or 'b' or both lacking a determinate denotation (as can happen for a vague definite description, e.g. 'the orange chair', or a name whose denotation is 'fixed' by such a description). In what follows I show how this possibility blocks the derivation of the contradiction by which Evans refutes the idea of vague objects, and then consider some difficulties, on this way of accounting for vague identity statements, for the view that *personal* identity might be vague.

I

Evans' derivation is as follows. Allowing for the expression of the idea of indeterminacy by the sentential operator '∇' we have

(1) $\nabla(a = b)$.

(1) reports a fact about b which we may express by ascribing to it the property '$\hat{x}[\nabla(x = a)]$':

(2) $\hat{x}[\nabla(x = a)]\,b$.

But we have:

(3) $\sim\nabla(a = a)$

and hence

(4) $\sim\hat{x}[\nabla(x = a)]\,a$.

But by Leibniz's Law we may derive from (2) and (4):

(5) $\sim(a = b)$

contradicting the assumption, with which we began, that the identity statement 'a = b' is of indeterminate truth-value.

Suppose now that 'a = b' is indeterminate in truth-value because, though the denotation of 'b' is definite that of 'a' is not: there are, say, two objects x and y such that on some sharpenings of the predicates embedded in 'a' (or in the description by which the denotation of 'a' was fixed) 'a' denotes x, on others y, and on others nothing at all. Then (1) is true, and so is (2): for b is either x or y so that the predicate '(x = a)' applies to it on some sharpenings and not on others (notice that on a sharpening of *all* predicates in (2) – or in the descriptions which fix the denotations of the names in (2) – (2) comes out as false, for on *every* sharpening of the predicate in 'a' or the description fixing its denotation the predicate '$\nabla(x = a)$' has a null extension). (3) is true also but (4) is either false or indeterminate in truth-value. For both x and y satisfy the predicate '$\nabla(x = a)$' since both satisfy '(x = a)' on some sharpenings and not on others. Hence if there is no sharpening of the predicates in 'a' or the description fixing its denotation on which it lacks denotation, '$\hat{x}[\nabla(x = a)]\,a$' is true and (4) is false, while if there is such a sharpening, both '$\hat{x}[\nabla(x = a)]\,a$' and (4) are indeterminate in truth-value. (Of course (4) is true under any sharpening of *all* the predicates occurring in it (or in the descriptions fixing the denotations of the names in it), but this no more makes it true than the fact, already noted, that (2) is false under any sharpening of all the predicates occurring in it makes *it* false.)

Suppose on the other hand that 'a = b' is indeterminate in truth-value because the denotation of 'b' is indeterminate though that of 'a' is not. Then (2) is false since the predicate '(x = a)' is sharply defined and the predicate '$\nabla(x = a)$' has a null extension.

Let us take it, then, that if an identity statement lacks a determinate truth-value this must be due to one or both of the singular terms flanking the identity sign lacking a determinate denotation.

II

Consider now the idea that in some of the puzzle cases described in the literature on personal identity the relevant identity statement is indeterminate in truth-value, e.g. the idea that 'Brownson = Brown' is indeterminate in truth-value. If this is so, either 'Brownson' or 'Brown' or both must lack a determinate denotation. Suppose Brown is the only man in room 101 before the operation and Brownson is the only man in room 101 after the operation, then 'the man in room 101 before the operation = the man in room 101 after the operation' must also lack a determinate truth-value, and one or both of the definite descriptions must lack a determinate denotation. Let us take 'the man in room 101 before the operation' as the candidate for lack of determinate denotation (which is chosen will make no difference to the argument). Suppose now that this has denotation just in case the predicate 'x is a man in room 101 before the operation and anything which is a man in room 101 before the operation is identical with x' is satisfied by just one object. Then if the denotation of the description is indeterminate it must either be (a) that some object x satisfies that predicate under some sharpenings and some object y under others or (b) that some object x satisfies it under some sharpenings and no object under others. For (b) to be so it will have to be the case that no, or several, objects satisfy 'x is a man in room 101 before the operation' under some sharpenings. The first possibility we can leave aside: it implies the absurdity that it is indeterminate whether there is a man in room 101 before the operation. The second possibility requires that there is more than one object which is a 'candidate' for satisfying 'x is a man in room 101 before the operation', i.e. satisfies it under some sharpenings, and this is also required by (a).

But now, how are these candidates to be identified? The philosopher who believes that persons are summations of temporal parts has an easy answer. One such candidate is the summation consisting of all Brown's pre-operation stages plus all Brownson's post-operation stages – for whether or not all the stages of this summation are linked by the personal unity relation it certainly *exists*. Another candidate is the summation consisting solely of all Brown's pre-operation stages – for it exists also, even if there are stages not contained in it linked by the personal unity relation to stages which are.

But what of the philosopher who believes that persons lack temporal parts? How can he specify the candidates? It is no use his giving a description of the form 'the person who did such-and-such and so-and-so at these times and those' – specifying details in the *pre-operation* history of Brown. For if the denotation of 'the man in room 101 before the operation' is indeterminate so, obviously enough, will be that of *any* such description. Nor will a description of the form 'the person who did such-and-such before the operation and so-and-so after the operation' be of any use. For such a description will have denotation just in case Brownson is Brown, and since that is *ex hypothesi* indeterminate it will also be indeterminate whether any such description in fact denotes.

I can think of no way at all, in fact, for the philosopher who does not believe that persons have temporal parts to respond to this request for specification, but even if he were able to do so his troubles would not be over. For the objects specified (the three-dimensional counterparts of the summations of thing-stages specified by the four-dimensionalist) would have to possess the following features: they would have to be (a) distinct, (b) both satisfiers of the predicate 'x is a man in room 101 before the operation' under some sharpenings, (c) located in exactly the same place before the operation and (d) not related either as part to whole or whole to part or as sharing parts. But it is hard to believe that there are any such objects. For suppose there are, and suppose our language becomes more precise in such a way as to render 'Brown = Brownson' false and 'Brownson = Robinson' true. Then one such object is the (now quite determinate) denotation of 'Brown', and another must be an object which it is now definitely false to describe as a person, but which coincides with Brown during his lifetime and with Brownson (Robinson) subsequently. But that the speakers of this more precise version of English must acknowledge the existence of such an object even if they do not think of persons as summations of temporal parts seems absurd, as absurd as my having to acknowledge the existence of an object which coincides with me throughout my lifetime and with my son subsequently but which is not merely a summation of temporal parts of myself and my son. I therefore conclude that if one takes the view that persons lack temporal parts one must either (a) deny the possibility of vague identity statements – or at least the possibility of vague statements of *personal identity* – as in fact appears to be done by Chisholm, or (b) deny that 'the man in room 101 before the operation' has denotation if and only if the predicate 'x is a man in room 101 before the operation and anything which is a man in room 101 before the operation is identical with x' is satisfied, and *mutatis mutandis* for other definite descriptions.

University of Birmingham

© HAROLD W. NOONAN 1982

SENTENCE MODIFIERS AND SEMANTIC THEORIES: A NOTE ON A CONJECTURE

By MARTIN DAVIES

IN *The Bounds of Sense* (Methuen 1966) Professor Strawson claims that the conclusion of Kant's Transcendental Deduction, semantics ('Operators, predicates and truth-theory', in Platts (ed.) *Reference, Truth and Reality*, Routledge & Kegan Paul, 1980). His conjecture concerns the treatment of expressions which syntactically are sentence modifers. (I restrict attention to one-place modifiers.)

In any semantic theory T for a language L there is, for each sentence s of L, a theorem of T stating the crucial semantic properties of s according to T. Such a theorem is the theory's *M-sentence* for s. The M-sentence for s might specify the meaning of s in L, it might specify material or strict truth (in L) conditions for s, and doubtless there are other possibilities. A sentence modifier O in L is said to be *treated as an operator* in a semantic theory T if O is governed by a proper axiom ($Ax\,O$) of T such that, for any sentence s of L, the M-sentence for $\ulcorner Os \urcorner$ is a deductive consequence of the M-sentence for s and $Ax\,O$. Thus paradigmatically, truth functional sentence modifiers, such as negation, are treated as operators in standard truth theories. On the other hand, some sentence modifiers such as 'John believes that' seem to resist treatment as operators in truth theories. The familiar paratactic account treats that modifier not as an operator but as a predicate (of utterances) plus a demonstrative singular term (referring to an utterance on an occasion of use).

The project of truth theoretic semantics, conceived in the broadest possible way, is the project of constructing semantic theories in which the concept of truth plays a fundamental role. It does not immediately follow that the concept of truth is itself employed in the M-sentences. Nor does it immediately follow that there are no theorems of the theory which specify the semantic properties of sentences more finely than do the M-sentences. Nevertheless, I shall make two assumptions about truth theoretic semantics which, although they may beg deep questions about what makes the concept of truth fundamental in a semantic theory, seem allowable in the present context. (i) I assume that the M-sentence for s is either of the form

s is true iff p

(in which case the M-sentence specifies material truth conditions for s), or else of the form

N (s is true iff p)

for some modifier N. (If N is '$\Box$', expressing broadly logical necessity, then the M-sentence specifies strict truth conditions for s.) (ii) I assume that, if what the M-sentence for s states does not distinguish s from some other sentence s', then there is no other theorem of the theory which specifies semantic properties of s which do distinguish s from s'. Thus, for example, I assume that if the M-sentence for s specifies material truth conditions for s (and so does not distinguish s from any materially equivalent sentence s'), then there is no other theorem of the theory which specifies strict truth conditions for s (and so does distinguish s from at least some materially equivalent sentences).

Does the distinction between those modifiers that permit, and those modifiers that resist, an operator treatment in truth theories correlate with any other interesting semantic distinction? McGinn notes a pattern in a range of examples

> an operator treatment has succeeded if and only if the construction in question produces a transparent context for (genuine) singular terms; it has failed if and only if the produced context is referentially opaque. (p. 202)

and conjectures that the pattern is quite general

> the transparency of a modifier and the feasibility of an operator treatment for it appear to coincide. (p. 203)

What I want to point out in this note is that McGinn's conjecture is not, in fact, correct.

What is and what is not possible within the project of truth theoretic semantics is quite sensitive to decisions about the resources that are to be allowed in truth theories. Thus suppose for a moment (what McGinn does not suppose) that it is required that truth theories be extensional, and that the M-sentences specify just material truth conditions. (These are features of 'standard' truth theories.) Then, suppose that the sentence modifier O does not admit substitution of material equivalents within its scope *salva veritate* (equivalently: O is non-extensional). This is to say that the material truth conditions of $\ulcorner Os \urcorner$ are sensitive to more than just the material truth conditions of s. If a truth theory is to respect this sensitivity, then no M-sentence for $\ulcorner Os \urcorner$ can be a deductive consequence of $Ax\,O$ together with the M-sentence for s. Otherwise the truth theory (including $Ax\,O$) will wrongly certify as material equivalents all pairs of sentences $\ulcorner Os \urcorner$ and $\ulcorner Os' \urcorner$ for which s and s' are materially equivalent. Thus, if it is required that truth theories be extensional, and that the M-sentences specify just material truth conditions, then the referentially transparent modal modifier '$\Box$' (expressing broadly logical necessity) will resist treatment as an operator. (It is true that this resistance can be broken down if rules of proof are allowed, for which the relevant notion of proof is proof from the proper axioms of the theory. But it seems to me that such rules of proof do not serve well the purposes of the semantic theorist, and, in any case, unrestricted use of such rules of proof will bring even referentially opaque modifiers within the scope of an otherwise extensional truth theory.)

The last paragraph provided an example of the way in which, quite generally, whether a modifier can be treated as an operator is sensitive to decisions about the resources that are to be allowed in truth theories. It is illuminating to divide truth theories into two kinds: (a) those whose M-sentences specify, at most, strict truth conditions for sentences (that is, broadly logically necessary and sufficient conditions for truth); (b) those whose M-sentences specify more than just strict truth conditions for sentences. I shall suppose first that truth theories are restricted to kind (a), and consider second whether the conjecture is correct if truth theories of kind (b) are allowed as well.

Suppose that truth theories are required to be of kind (a). Then, suppose that the sentence modifer O does not admit substitution of strict equivalents (that is, necessary equivalents, where the necessity in question is broadly logical necessity) within its scope *salva veritate* (equivalently: O is hyperintensional). This is to say that the material and strict truth conditions of $\ulcorner Os \urcorner$ are sensitive to more than just the strict truth conditions of s. If a truth theory is to respect this sensitivity, then no M-sentence for $\ulcorner Os \urcorner$ can be a deductive consequence of $Ax\, O$ together with the M-sentence for s. Otherwise the truth theory (including $Ax\, O$) will wrongly certify as strict equivalents all pairs of sentences $\ulcorner Os \urcorner$ and $\ulcorner Os' \urcorner$ for which s and s' are strictly equivalent. Thus, if it is required that truth theories be of kind (a), then any referentially opaque modifier will resist treatment as an operator. (Since two sentences differing only in which of two co-referring genuine singular terms is used are strictly equivalent (that is, broadly logically necessarily equivalent), a referentially opaque modifier does not admit substitution of strict equivalents *salva veritate*.) So far so good for the conjecture. But some referentially transparent modifiers will also resist treatment as operators, namely any sentence modifiers that are hyperintensional but referentially transparent. The modifier 'John believes that' is, according to some theorists, such a modifier (e.g. Barwise and Perry, 'Semantic innocence and uncompromising situations', in French *et al.* (eds), *Midwest Studies in Philosophy*, Volume VI, 1981). But, in any case, a sentence modifier can be introduced with a stipulation that it is referentially transparent. And the possibility of the introduction of such a hyperintensional modifier is enough to show that the conjecture is not correct if truth theories are required to be of kind (a).

Suppose, on the other hand, that truth theories of kind (b) are allowed as well. In this case, the problem for the conjecture is not that referentially transparent modifiers resist operator treatments, but that referentially opaque modifiers admit of operator treatments. John Wallace has offered a way of constructing truth theories of kind (b), by prefixing the familiar axioms and theorems of a standard truth theory with the modifier 'it is a matter of meaning alone that' ('Logical form, meaning, translation', in Guenthner and Guenthner-Reutter (eds), *Meaning and Translation*, Duckworth 1978). Clearly enough, an M-sentence of the form

> it is a matter of meaning alone that s is true iff p

does more than specify strict truth conditions for the sentence s. In such a theory, the modifier 'John believes that' cannot be governed by any principle that has the following (initially plausible) consequence.

> if (it is a matter of meaning alone that s is true iff p) then (it is a matter of meaning alone that ⌜John believes that s⌝ is true iff John believes that p).

For if s means, say, that grass is green, then along with

> it is a matter of meaning alone that ⌜$\sim s$⌝ is true iff $\sim$(grass is green)

we shall have in a theory of the kind offered by Wallace

> it is a matter of meaning alone that ⌜$\sim s$⌝ is true iff $\sim$(s is true).

Yet we do not want

> it is a matter of meaning alone that ⌜John believes that $\sim s$⌝ is true iff John believes that $\sim$(s is true).

The complication proposed by Wallace is that we should define

> s means that p

using the two modifiers 'it is a matter of meaning alone that' and 'it is a matter of the laws of meaning applied in the left-right direction only that' as follows.

> It is a matter of meaning alone that s is true iff p and for all q, if it is a matter of meaning alone that s is true iff q, then it is a matter of the laws of meaning applied in the left-right direction only that if q, then p. (p. 50)

Then the modifier 'John believes that' can be governed by the principle

> it is a matter of meaning alone that, for all sentences s, ⌜John believes that s⌝ is true iff John believes what s means. (ibid.)

If such a truth theory of kind (b) is allowed, then hyperintensional modifiers, be they ever so referentially opaque, can be treated as operators.[1] (Wallace's suggested theory is in some ways analogous to a suggestion of Baldwin's discussed by McGinn. What led to the complication proposed by Wallace is analogous to the first of two objections made by McGinn against Baldwin (p. 203). Nothing analogous to McGinn's second objection seems to hold against Wallace.)

---

[1] Perhaps theorems whose abbreviations are of the form

> s means that p

should be regarded as M-sentences.

I am not suggesting that the semantic theorist does well to consider truth theories of kind (b). On the contrary, it seems most natural to restrict truth theories to kind (a). If the semantic theorist wants a theory that does more than specify (broadly logically) necessary and sufficient conditions for the truth of sentences, then he can have what he wants by constructing a theory that explicitly invokes the concept of meaning and straightforwardly specifies what sentences mean. What lies behind the availability of a truth theory of the kind Wallace offers is, it seems to me, the availability of a semantic theory that is not a truth theory at all, but a meaning theory (strictly so-called). And what lies behind the incorrectness of the conjecture if truth theories of kind (b) are allowed is, it seems to me, the fact that hyperintensional sentence modifiers can be treated as operators in a meaning theory (strictly so-called), a fact which itself goes against the spirit of the conjecture.

If truth theories are restricted to kind (a), then we can say that just as non-extensionality blocks operator treatments in extensional truth theories, so hyperintensionality blocks operator treatments in truth theories *simpliciter.* That last fact does not, by itself, warrant the conclusion that hyperintensional modifiers are, in semantic reality, not operators but predicates. Rather it points to the two (not very exciting) facts (i) that some sentence modifiers form complex sentences whose truth conditions are sensitive to the meaning, and not just the truth conditions, of the contained sentence; and (ii) that a truth theory does not specify what sentences mean.

Although the conjecture is incorrect, there is undoubtedly something right about the motivation behind it. For it is epistemological, rather than modal, modifiers that introduce hyperintensionality (which blocks operator treatments in truth theories). It is in the field of epistemology, rather than modality, that we must consider not just singular reference to objects but, in addition, different ways of thinking about objects. And arguably it is different ways of thinking about objects that are ultimately responsible for (apparent?) failures of substitutivity of co-referring genuine singular terms *salva veritate.*[2]

Birkbeck College, London

© MARTIN DAVIES 1982

---

[2] Lloyd Humberstone and Barry Taylor made helpful comments on an earlier version of this note.

TRANSLATION AND BELIEF

By ARNOLD CUSMARIU

ALONZO CHURCH gave an argument some years ago (see 'On Carnap's Analysis of Assertion and Belief', ANALYSIS 10.5, April 1950, pp. 97–9) intended to refute the theory of belief Carnap had proposed in *Meaning and Necessity* (University of Chicago Press, 1947, pp. 54–5, 62). The argument has engendered a great deal of controversy.[1] No consensus has been reached, however, as to what the argument can minimally assume in order to succeed. I will suggest below a streamlined version and show that it bypasses some well-known objections.

I

According to Carnap,

> A. Smith believes that snow is white

is to be analysed as

> B. There is a sentence Z in a semantical system S′ such that (a) Z in S′ is intensionally isomorphic to 'Snow is white' in English, and (b) Smith is disposed to an affirmative response to Z as a sentence of S′ (p. 62).

Carnap rejects as inequivalent with A simpler versions of B such as

> C. Smith is disposed to respond affirmatively to some sentence in some language, which is logically equivalent with the English sentence 'Snow is white' (p. 55),

and

> D. Smith is disposed to an affirmative response to 'Snow is white' as an English sentence (pp. 54–5);

he even grants that B might not be the final analysis (p. 62).

Unlike other objections that have been raised against Carnap's analysis, Church's argument is not aimed at its behaviourist unpacking of 'believes'. For purposes of stating this argument we may therefore take as the analysis of A:

> E. There is a sentence Z in a semantical system S′ such that

---

[1] Here is a partial list. R. Carnap, 'On Belief Sentences: Reply to Alonzo Church', in M. Macdonald (ed.), *Philosophy and Analysis* (Blackwell, 1954), pp. 128–31. W. V. Quine, *Word and Object* (MIT Press, 1960), pp. 213–14. R. M. Martin, *Belief, Existence and Meaning* (New York University Press, 1969), pp. 61–4. R. J. Haack, 'Translation, Analysis and Ontology', *Review of Metaphysics* 27 (1973), pp. 298–317. M. Dummett, *Frege: Philosophy of Language* (Harper & Row, 1973), pp. 370–3. S. Leeds, 'Church's Translation Argument', *Canadian Journal of Philosophy* 9 (1979), pp. 43–51.

(a) Z in S′ is intensionally isomorphic to 'Snow is white' in English, and (b) Smith believes Z as a sentence of S′,

with the understanding that B and E stand or fall together.

II

Church's argument that E is not a correct analysis of A is best understood in two stages. First, consider translations of A and E into, for example, German:

> A′. Smith glaubt, dass der Schnee weiss ist.
> E′. Es gibt einen Satz auf einem semantischen System S′ so dass (a) Z als Satz von S′ intensional isomorph zu 'Snow is white' als Englischer Satz ist, und (b) Smith Z als Satz von S′ glaubt.

A was easy to translate. Most of E is also unproblematic, except for the end of clause (a). It seems there are three other ways the translation might have gone:

> E1. . . . 'Snow is white' als Deutscher Satz ist . . .
> E2. . . . 'Der Schnee ist weiss' als Englischer Satz ist . . .
> E3. . . . 'Der Schnee ist weiss' als Deutscher Satz ist . . .

The question is, why is E′ the correct translation of E and not any of these? The first two possibilities are easy enough to rule out. E1 and E2 are incorrect translations because, unlike E, they are clearly false. 'Snow is white' is not a sentence of German as E1 implies, nor is 'Der Schnee ist weiss' a sentence of English as E2 implies. Correct translation ought surely to be truth-preserving.

Ruling out E3 is not quite this simple, however. As I read Church, he would fault E3 on grounds that it does not contain the same information as E. This gives us the following principle:

> P. Sentence X′ of natural language L′ is a correct translation of sentence X of natural language L only if X′ contains the same information as X.

The requirement here is reasonable enough, only it brings up the difficult problem of what constitutes sameness of information. Church himself thinks that sameness of information must be preserved by a correct analysis as well (pp. 97-8), but does not pause to make the requirement completely precise. As it happens, this difficult problem need not be solved here; a necessary condition will do:

> P*. Sentence X of natural language L contains the same information as sentence X′ of natural language L′ only if X implies whatever X′ implies.

Now, E3 implies that Z is intensionally isomorphic with the German sentence 'Der Schnee ist weiss' but it is clear that E does not. It follows that E and E3 do not contain the same information, by P*. It follows further that E3 is not a correct translation of E into German, by P. Assuming such a translation is to be culled from among E′, E1, E2 and E3, we have an argument for the conclusion that E′ is the correct translation of E into German.

This completes the first stage of Church's argument.

At the second stage we find the point that A′ and E′ would convey different meanings to a German who happened to be ignorant of English (p. 98), whereas A and E would not pose a problem for English users ignorant of German. From these observations Church concludes that E is not an adequate analysis of A.

The argument at this stage is an enthymeme. To complete it, we need to make explicit the connection between translation and analysis, on the one hand, and the observation that an epistemic disanalogy exists between German and English users as regards access to Carnap's analysis. The following principle makes this connection explicit:

> Q. If Y is a correct analysis of X as expressed in language L, and if X′ and Y′ are correct translations of X and Y into language L′, and if users of L ignorant of L′ are able to fully understand X and Y, then users of L′ ignorant of L must also be able to fully understand X′ and Y′.

Now, it seems clear that Germans ignorant of English would not be able to fully understand E′, for the reason that they would not be in a position to understand the sentence quoted therein. It follows either that E is not a correct analysis of A as expressed in English, or that English users ignorant of German could not fully understand both A and E, or that A′ and E′ are not correct translations of A and E into German. The last of these is ruled out by the first stage of Church's argument. The middle possibility is completely implausible. Therefore, E is not a correct analysis of A as expressed in English – or any other language for that matter, since the choice of language has been arbitrary.

Q seems intuitively plausible. It is merely a principle of epistemic parity reflecting the idea that, while knowledge of languages may be a useful skill, it ought not be required in one case when it is clearly not required in another case which is relevantly similar. Users of only one language have a right to expect equal access to an analysis translated correctly from another language. (Parity is also the reason why understanding the sentence 'Snow is white' quoted in E′ is a necessary condition of fully understanding E′. For English users who failed to understand this sentence would not be said to fully understand E.)

III

In the various discussions of Church's argument one finds more ambitious principles than P and Q attributed to him. The situation is perhaps understandable since Church himself is not entirely explicit about what principles he needs to accept to have a valid argument. He speaks as if he would accept also these:

R. X′ is a correct translation of X from L into L′ only if X′ as a sentence of L′ is synonymous with X as a sentence of L.

S. If Y is a correct analysis of X, then Y contains the same information as X.

T. If Y is a correct analysis of X, then Y reproduces the meaning of X.

W. V. Quine finds Church's argument inconclusive owing to T's appeal to the notion of sameness of meaning (p. 214). His reasons for doubting the intelligibility of this notion are too well known to rehearse here. (Those reasons would also affect R.)

R. M. Martin finds requirement S excessive. He makes the point that a simple arithmetical formula such as '1 + 1 = 2' does not contain the same information as its complex counterpart in *Principia Mathematica*. But from this it doesn't follow that the *PM* version is an incorrect analysis of this formula (p. 63).

Last, M. Dummett objects that principle R is too strong:

> There is no ground for presumption that the practical canons of apt translation always require strict synonymy (p. 372).

Take the last objection first. Dummett could be right about practical canons and wrong about philosophical ones. The fact that synonymy is not always required in fiction or historical narrative – which he mentions – does not show that the same is true in philosophy or other contexts (diplomacy, law) where greater rigour is the norm. Clearly, Church need not assume that strict standards of translation can be applied to every context. Anyhow, we need not decide the fate of synonymy here. I have used a weaker notion of translation which requires not sameness of meaning but the next best thing, sameness of consequences. It is reasonable to employ so strict a notion in philosophy where others (identity, analysis) are also used strictly, even if a looser one would be more apt in less rigorous contexts. Thus Dummett's objection is bypassed.

Martin may be right that, in some sense (which he does not specify), '1 + 1 = 2' and its *PM* version do not contain the same information. But the objection has no force against P since this principle is about translation, not analysis, as applied to natural languages, not formal ones. (Also, I am sceptical that a purely extensional notion of information would yield Martin's point.)

Quine's misgivings over synonymy may well be justified, but they do not affect the intelligibility of P and Q. My epistemic reading of Church's phrase 'conveys different meanings', as incorporated in principle Q, clearly avoids appeal to the notion of sameness of meaning; likewise the extensional reading given in P* to the notion of sameness of information.

Lest these replies make me seem uncharitable, let me conclude by granting that if Church were to need R, S or T to argue successfully against Carnap's analysis of belief, then the above objections would be very serious stumbling blocks indeed. What I have tried to show in this paper is that there exists a version of his argument which can succeed with much less.[2]

© ARNOLD CUSMARIU 1982

---

[2] I have been helped by comments from Herbert Heidelberger, Michael Levin, Philip Quinn and from Professor Church.

THE CONCEPT OF EXPERIENCE AND STRAWSON'S TRANSCENDENTAL DEDUCTION

By KIM DAVIES

IN *The Bounds of Sense* (Methuen 1966) Professor Strawson claims that the conclusion of Kant's Transcendental Deduction, namely that the contents of experience must be brought under concepts of an objective world, represents a 'very great and novel' gain in epistemology (p. 29). I wish to argue that closer consideration of the concept of experience with which Strawson, following Kant, begins, pre-empts the whole project.

Strawson sets out the following thesis assumed as a premise in (his reconstruction of) the Transcendental Deduction:

> that there must be such unity among the members of some temporally extended series of experiences as is required for the possibility of self-consciousness, or self-ascription of experiences, on the part of a subject of such experiences (the thesis of the necessary unity of consciousness) (p. 24).

This thesis is linked with the dual character of experience (viz. that *particular* contents of experience should be recognized as having some *general* character) and, as so linked, is taken as a 'standard-setting definition of what is to count as "experience"' (p. 25). Why should we begin with this definition? Strawson notes that other forms of sentience, short of this standard, may exist, but goes on to argue not only that no other philosopher 'even the most economical of empiricists' has tried to work with a more limited conception (p. 25), but that only this concept of experience 'can be of interest to us' (p. 28-9). It is not clear to me that writers such as Hume, Russell and Ayer do in fact build the possibility of self-consciousness into their concept of experience, but I wish to discuss the question of the interest which binds together the community indicated by Strawson's first-person plural.

The interest which a concept of experience has for 'us' is presumably to be understood in the light of the purposes 'we' share. Of course, we do in fact self-ascribe experiences, but dogs and babies do not, and it might be thought that this less sophisticated form of sentience was of more interest. To see why it is not, we should ask why the concept of experience is important for philosophy. It seems to me clear that its fundamental significance lies in its opening a gap, between how things are experienced as being and how they are, which is the ground of the possibility of any critical philosophy. In the light of the possible disparity between how things are and how they are experienced as being, the question of the justification of claims to knowledge emerges. Without the recognition of this possible disparity, this epistemological gap, the interrogation of knowledge-claims which is central to all philosophy (other than the 'angels on pins' type) would not be possible. It underlies the Cartesian Doubt and those traditions in Western Philosophy which followed in its wake. Now it would seem that for someone to possess the concept of experience required here, and to recognize the epistemological gap, the self-ascription of experience is required – it looked as though X, but in fact Y – or at least that self-reflexiveness of experience which for Strawson constitutes the core of empirical self-consciousness (p. 111). Thus can we understand the philosophical interest of a concept of experience into which the possibility of self-ascription is built. And if experience is to provide a kind of alternative to successful cognition of an objective world, on the basis of which we can call for the justification of knowledge-claims, then it must have the dual character Strawson talks of, at least insofar as it must comprise some kind of awareness or consciousness of some kind of item, which recognizes that item as falling under a general concept. Thus can we understand the philosophical interest of the second component of Strawson's concept of experience, so that it may seem that we must agree with him that his definition as a whole is 'surely acceptable'.

Let us first look closer at the 'object' side of this concept. Strawson's word is 'accusative', indicating that it is internal to the experience, and not necessarily (at this stage) conceived of as independent of experience. If it were so conceived, then the claim that concepts of the objective must be used in experience would be built into the concept of experience from the start, so that an argument to that effect would hardly be required. But this is precisely what must be the case if the concept of experience is to be adequate to the philosophical purposes it is to serve. For unless experience is of items conceived of as independent of experience itself, there would be no possibility of opening up a gap between how things are and how they are experienced as being, and so no possibility of critical philosophical thought. If experience involved nothing more than an awareness of an item which made no claim to be independent of experience, for example the red patches and tickling sensations of the sense-datum theorists (p. 99), then there could be no disparity between how things are experienced as being and how they are, since there would be nothing more to how things are than how they are experienced as being. That experience is of items conceived of as existing independently of experience itself, that those items are brought under 'concepts of the objective', must already be built in to that concept of experience which is to be of fundamental interest to us. Without this as the primary concept of experience, Kant could never have faced a scandal, nor Strawson his sceptic. Strawson's reconstruction of the Transcendental Deduction can succeed only in reaching the point which it first started from.

Although this account retains the value of some of Kant's Analytic, particularly the Analogies which lay out the necessary structure of the conception of an objective world in terms of substance and causality, and although the kind of considerations prompted by Kant's arguments have a value which transcends this argument itself, the fact that it takes a major thread of the Deduction to be, at bottom, idle, might call for more discussion. How is it possible that Kant, and then Strawson, could have failed to recognize this? The answer lies in a forgetfulness of the provenance of their concept of experience inscribed in their work. We can see how, on the basis of a concept of experience as primarily of items brought under concepts of the objective, we could recognize that some experiences do not have this character, and then wish to generalize our concept to that of an awareness of an item, where nothing is presupposed about the nature of that item. It is not then difficult to focus, in the spirit of scientific generality, on this concept at the cost of a gradual amnesia concerning the circumstances in which it emerged, and the features needed for the primary concept to serve its philosophical purpose. This is effectively what occurred between the time of Descartes and Hume; Kant (with Strawson following) takes up Hume's problems and with them the concept of experience which they presuppose. Now there is certainly some intrinsic interest in showing that if one starts with this concept, then the contents of experience must be brought, at least in part, under concepts of the objective. It nevertheless remains the case that this requirement must be built into any concept of experience which, as the ground of the possibility of the demand for justification, can be of fundamental significance to critical philosophy, and so explain the philosophical interest which Strawson speaks vaguely of. That the Transcendental Deduction can be held to have secured very great and novel gains is thus a mark of the thoroughness with which critical philosophy forgets its beginnings, and of the need for the kind of philosophical self-reflection attempted here.

University of Durham © Kim Davies 1982

MACKIE'S DEFENCE OF INDUCTION

By P. J. R. Millican

I

HUME's attack on induction is very familiar: *that the past be a rule for the future* cannot be established deductively, whilst any inductive argument to that conclusion would itself depend on it, and so be circular. But Hume takes for granted that there are no other kinds of reasoning available besides deduction and induction, and therefore assumes that the possibilities for justification of the latter are exhausted, since neither of them is adequate to the task. If his rigid dichotomy is accepted, Hume's argument indeed seems compelling. Recently, however, it has been challenged by those who see a third possibility, and I would like to consider, as one representative of this approach, J. L. Mackie's attempt to provide a *probabilistic* justification of induction.[1]

Mackie sets himself the task of justifying a prediction of *general* uniformity for a limited period, a prediction which can then itself be used to support other, particular inductions. He considers first the simple pair of rival hypotheses, that the world's ways of working are completely uniform throughout, and that they are completely random. 'If we had just these alternatives to choose between, it would be reasonable to prefer the former in the light of our observations, unless it was antecedently almost infinitely less probable than the second.' And to assume that it was so vastly improbable would, of course, be question-begging.

Unfortunately, the matter is complicated by the profusion of other possible hypotheses, even if we accept it to be overwhelmingly probable that the world has in fact been completely uniform during the period of our past observations. For it could be that this uniformity has a limited temporal (or spatial) range, beyond which it terminates or gradually fades out. These hypotheses of extensive but limited order have yet to be dealt with, and Mackie's central argument is directed to this purpose:

---

[1] J. L. Mackie, 'A Defence of Induction', in G. F. Macdonald (ed.), *Perception and Identity* (London and Basingstoke: Macmillan, 1979).

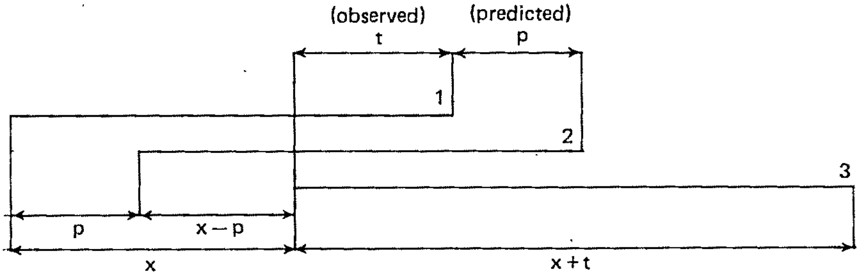
> First, let us consider the indefinitely large set of hypotheses all of which assign the same spatio-temporal extent to uniformity, but locate it differently.... It seems a legitimate application of the principle of indifference to assign equal antecedent probabilities to the various hypotheses of this set. Some of them, however, will have been ruled out by the already-observed spread of uniformity.... Then, of the hypotheses not so ruled out, if the extent which is common to the hypotheses of the set... is considerably greater than the observed spread of uniformity, relatively few will say that uniformity will terminate either at once or very soon. So,... we can conclude that, even if uniformity lasts only for this limited extent, it is not likely to end very soon. But, secondly, let us consider the set of sets of such hypotheses.... Clearly, the shorter the extent characteristic of each set of hypotheses, the smaller the proportion of the hypotheses of that set that will cover the observed spread of uniformity.... Consequently, by an inverse probability argument, the observation of a certain spread of uniformity raises the probability that the extent of uniformity is considerably greater than that spread much more than it raises the probability that that extent is equal to or only a little greater than that spread. Now,... we can appeal to a... principle of tolerance to justify our not giving a zero antecedent probability to all hypotheses assigning more than a certain extent to uniformity. So long as the greater-extent hypotheses are not ruled out by such an unfair initial assignment, they can come out more probable in the end, their probability being raised more by the observation of some considerable spread of uniformity. And, as we have seen, once we have confirmed such a greater-extent hypothesis, we can go on by a direct probablity argument to infer that the uniformity is not likely to end very soon.

All this seems very plausible, until we examine it mathematically.

Let us assume that one of Mackie's hypotheses of limited order is correct, and that we have already observed uniformity for a length of time t, predicting that it will continue for a further time p. Now, if P(x) is the probability density function representing the initial probability that uniformity will extend over a total time x, then since, by the principle of indifference, all hypotheses assigning an equal extent to uniformity are equally probable, it follows that the probability that uniformity is of duration x and begins at some point within a certain interval of time i is proportional both to P(x) and to i, say kiP(x), where k is some constant. With this information, we want to find the probability that uniformity will continue for at least a further time p, fulfilling our prediction to that effect.

Consider the set of hypotheses whose characteristic extent $(x + t)$ exceeds the observed spread of uniformity t by a time x $(x > p)$:

105 An 13



Lines 1, 2 and 3 represent three possible uniformities of duration (x + t), all of which account for the observed uniformity t. 1 ends immediately, 2 continues for a further time p, while 3, which began at the time of our first observation, finishes after time x. There is, then, a range of possible uniformities of the given duration (x + t) which are consistent with our observations: they all start at different points within the marked period x, and of these, the uniformities starting within the marked period (x – p) will yield successful predictions. As Mackie says, the proportion (x – p)/x of successful predictions will grow considerably as x is increased. In addition – Mackie's second point – a greater value of x will bring a greater range x of possible (starting points for) uniformities, and we can therefore use an inverse probability argument to confirm those hypotheses involving a large value of x, at the expense of those involving a smaller value.

For any *particular* value of x, the initial probability of a uniformity of extent (x + t) which will cover the observed spread of uniformity t is kxP(x + t). So the initial probability of a uniformity of *any* extent which will cover this spread is equal to the sum of the function kxP(x + t) for *all possible values of* x from zero to infinity (let us write this as 'SUM_0(kxP(x + t))', indicating by the '0' the value of the lower limit of x). In this sum, the multiplication by x will give greater weight to kP(x + t) for larger values of x, and this is the basis of Mackie's inverse probability argument. But it should be noted that this weighting is quite independent of t – no matter how much uniformity is observed, the probability that total uniformity will last for a period which exceeds observed uniformity by an excess time x is raised, on account of that observation, in proportion to x alone.

Let us now move on from Mackie's inverse probability argument to his direct probability argument. As we have seen, if total uniformity exceeds observed uniformity by a time x, then the probability of a successful prediction to the effect that uniformity will continue for a time $p (x > p)$ is (x – p)/x. But again, this value is completely independent of t – the probability in no way improves as our observation of uniformity continues. So neither part of Mackie's argument can justify at all taking the past as a rule for the future, for if the past is to be a rule for the future, then, at the very least, a greater observed regularity must give greater strength to a constant prediction. And this argument makes no mention of the extent of the observed regularity: 't' appears only in the argument-place of the function P(x), which Mackie does not discuss.

If induction is to be justified by some method such as Mackie's, this must depend entirely on the behaviour of the function P(x). We can establish the condition for such a justification using Bayes' theorem:

$$\text{Probability of event, given evidence} = \frac{\text{Initial probability of both}}{\text{Initial probability of evidence}}.$$

Our 'evidence' is the observed uniformity t, while the predicted 'event' is a spread of uniformity covering *both* the observed time t *and* the predicted time p. In this case, then, the initial probability of *both* evidence and event is equal to the initial probability of the event, since any example of the latter must be an example of the former as well.

The initial probability of the observed uniformity t is

$$\mathrm{SUM}_0(\mathrm{kxP}(x+t)).$$

By an exactly parallel argument, the initial probability of the observed *and* predicted uniformity (t + p) is $\mathrm{SUM}_0(\mathrm{kxP}(x+t+p))$. Applying Bayes' theorem, we reach the following result:

Probability of a successful prediction

$$= \mathrm{SUM}_0(\mathrm{kxP}(x+t-p))/\mathrm{SUM}_0(\mathrm{kxP}(x+t))$$

$$= \mathrm{SUM}_0(\mathrm{xP}(x+t+p))/\mathrm{SUM}_0(\mathrm{xP}(x+t)).$$

To justify taking the past as a rule for the future, this probability must grow as t increases. This condition will be satisfied if P(x + p)/P(x) is an increasing function[2]; but this is hardly surprising, since that is the same condition required by Bayes' theorem if it is to be the case that the longer a regularity goes on, the greater the chance of its continuing for a further time p.[3] Clearly, to claim that this is so is to beg the very question at issue: to assume that the past *is* a rule for the future.

---

[2] Since in that case the functions P(x + t + p) and P(x + t) provide a more favourable ratio as t gets larger while at the same time the minimum arguments for which they are evaluated – i.e. when x = 0, giving minimum arguments of (t + p) and t respectively – increase, thus excluding those lower valued arguments which would provide a less favourable ratio.

[3] More precisely, Bayes' theorem yields the condition that $\mathrm{SUM}_{t+p}(P(x))/\mathrm{SUM}_t(P(x))$ be an increasing function of t, but this comes to the same thing.

II

Mackie's defence of induction fails because he considers only the two possibilities, that the past is *some* positive rule for the future, and that the world is completely random. He can then claim, quite reasonably, that a dogmatic assumption of randomness would be extreme and arbitrary, and that we must therefore 'allow from the start that there is some better-than-zero probability of some not purely random pattern'. In our terms, Mackie assumes that either $P(x+p)/P(x)$ is an increasing function (i.e. that the longer a regularity goes on, the longer it is likely to continue) or that it is a constant (i.e. that the past history of a regularity has no effect whatever on its future – his talk of a simple 'pure randomness' here is misleading, for *any* probability density function of the form $P(x) = a\,e^{-ax}$ will yield this result). A possibility which Mackie does not consider is that $P(x+p)/P(x)$ is a decreasing (or an oscillating) function, and his argument will not succeed unless he can rule this out. There are two ways in which he might attempt to do so: one of these is pragmatic, and the other logical.[4]

First, it might be maintained that we have good reason to proceed upon the assumption that $P(x+p)/P(x)$ is an increasing function, for only if this is so can we draw any conclusions about the future at all. We are unable to justify the assumption in any other way, but must simply adopt it on the principle that even a slight chance of success is better than the certainty of failure. Such a defence of induction, however, relies upon the fact that we do actually predict inductively, and cannot help doing so: it is for precisely this reason that our forecasts can be reliable only if $P(x)$ is such as to validate induction. But if we cannot help reasoning inductively, then any *pragmatic* justification of induction is quite unnecessary in the first place; certainly there is no need whatever to resort to anything as sophisticated as probability theory in order to persuade us to do what we cannot but do!

The second approach is to argue that there is something meta-inductively inconsistent about the suggestion that $P(x+p)/P(x)$ should be a decreasing function. For this would imply that the world was in some way counter-inductive, or at least that the longer uniformity lasts, the less likely it is to continue. Now it is clear that if in such a world inductions consistently fail, then nevertheless one *meta*-induction, to the effect that inductions will *continue* to fail, must be successful. Counter-induction cannot give a coherent direction on all levels simultaneously, and this is sufficient to rule it out as a comprehensive predictive policy.

This reply to the sceptic is open to two objections. First, it is of no assistance to Mackie, for it has force only against an all-embracing counter-induction, and none against a counter-induction which is applied specifically to predictions of general uniformity (Mackie's chosen battle-ground). Secondly, it makes the pragmatic assumption that even the sceptic must be committed to *some* comprehensive predictive policy, after which it need but show that induction is the most consistent such policy available. This assumption, however, is just what the critic of induction will deny. Why suppose that *any* method of forecasting should be reliable, in the absence of an independent proof that the past is a rule for the future?

---

[4] Both were suggested to me by Mr Mackie, for whose helpful discussion I am extremely grateful.

III

Mackie's argument fails for two reasons. First, even if a long observed uniformity increases the probability of a longer total uniformity, it does *not* follow that the probability of a certain length p of *unobserved* uniformity is also increased. For as the observed length t is increased, so also the predicted length $(t+p)$ is increased – the prediction is not constant, and is not in any way confirmed by Mackie's argument. Secondly, Mackie has no way of supplementing his argument with information about the function $P(x)$. Even if it is assumed that assignments of 'initial' probability (i.e. in advance of *all* experience!) about the future make sense; even if such problems as Goodman's paradox are overlooked; still there is nothing that can justify the particular assumption that the world's ways of working are inductive rather than counter-inductive. To put the point in a Humean fashion: no deductive argument can do the job, any inductive argument will beg the question, while a probabilistic argument will be circular, simply repeating the problems at a deeper level.

Lincoln College, Oxford

© Peter Millican 1982

CHISHOLM'S OBJECTION TO PHENOMENALISM

By MICHAEL J. MALONEY

IN the appendix to his book *Perceiving: A Philosophical Study*, Roderick Chisholm has argued that there is no physical object statement (POS) that entails any sense datum statement (SDS), and that, as a consequence, phenomenalism is false. This argument has come to be regarded as a classical refutation of phenomenalism, but I would like to argue in this paper that his objection to that position does not succeed.

Chisholm asks us to take any POS (P) and its alleged entailment (S), a sense datum statement. It is always possible, he argues, to come up with a statement (Q) such that Q is a 'thing statement referring to observation conditions' consistent with P and such that the conjunction of P and Q entails ~S. Hence P alone does not entail S.

That is,

> For any POS (P) alleged to entail a SDS (S) there is an observation condition statement (Q) consistent with P such that P & Q entails ~ S.
>
> Therefore it is not the case that any POS entails a SDS.

Chisholm grants that the conjunction of P and some other thing statement Q′, referring to observation conditions, does entail S. But let P′ = (P & Q′). This, on Chisholm's account, does entail a SDS (and so there is no statement consistent with it such that P′ and that statement entail ~ S). Is P′ a POS? We can accept for now some uncontroversial case of a POS in place of P (say, 'the table is brown'). What about Q′? Well, Q′ will be a 'thing statement referring to observation conditions' and where Chisholm gives examples of such statements they refer to such things as the lighting conditions and the condition of someone's sense organs. These are also physical object statements, and since we can assume that any conjunction of physical object statements is itself a physical object statement there are physical object statements that entail sense datum statements.

Let us examine another type of statement. An instance of this type of statement (P″) would be any conjunction including one POS (P) and the negations of any observation condition statements (Q) such that Q is consistent with P and the conjunction of P and Q entails ~S. (That is, the *negation* of any observation condition statement that Chisholm would conjoin with P to show that it does not entail the relevant SDS, (S), is to be conjoined with P). It is, epistemologically, an open question whether P″ entails S, and there is no observation condition statement Q such that it is consistent with P″ and (P″ & Q) entails ~ S. Hence no Chisholmian counterexamples can be constructed to defeat the claim that P″ does indeed entail S and his objection fails for statements of this type as well. (Of course P″ might entail neither S nor ~ S; but the claim of the phenomenalist is protected against Chisholm style counter-examples).

I return to the question of whether P′ and P″ are physical object statements. I assume that observation condition statements are all statements about physical objects (states of affairs etc.). But even if Q is about such things it is not obvious that ~ Q is as well. (Plantinga has urged that statements like 'It is not the case that Socrates is snubnosed' are not about Socrates, but about the proposition that Socrates is snubnosed.) What is missing from Chisholm's objection is an account of physical object, observation condition, and sense datum statements. Without such an account there is no way to avoid such objections as those advanced above. I will finish with some remarks displaying some of the difficulties that will be encountered in the effort to come up with an account of physical object statements sufficient for Chisholm's purposes.

It will certainly not do to say that P is a physical object statement iff P is about physical objects. (P & S), where S is a sense datum statement, would then be a physical object statement that entails a sense datum statement. Nor will it do to say that P is a physical object statement iff P is about physical objects and only about physical objects. Aside from the possibility that the claim would make Chisholm's objection trivially true, it will not work at least if we accept Chisholm's account (in *Thought and Its Reference*) of what it is for a proposition to be 'about' something. It is a consequence of that account that no proposition is only about physical objects. His position is that a proposition P is about a particular thing x if P implies x to have some property. His definition of 'P implies x to have the property F' is:

(D4) There is a property G such that (i) only one thing can have G at a time, (ii) P entails the conjunction of G and the property F, and (iii) x has G.

Substituting for x the proposition P itself, and for F the property of P's being true, we find that there is a property G (the property of being the proposition that P) such that (i) only one thing can have G at a time, (ii) P entails the conjunction of G and the property F (that is, P entails a proposition, itself, which is necessarily such that it is true iff something has the property of P's being true), and (iii) P has G. Therefore every proposition is about itself and so no proposition is about physical objects and only physical objects. Taking Plantinga's suggestion to heart, P″ will be about physical objects and propositions. But so will any physical object proposition if Chisholm is right. If any proposition is a 'pure' physical object proposition, then there is no reason to deny that status to P″.

Syracuse University © Michael J. Maloney 1982

THE PARADOX OF THE 1,001 CATS

By E. J. LOWE

IN §110 of his book *Reference and Generality* (Third Edition; Cornell University Press, 1980), P. T. Geach presents the following puzzle or paradox. We are to suppose that a certain cat, Tibbles, is sitting on a mat; moreover, Tibbles is the only cat sitting on the mat. Since Tibbles is, we suppose, a normal cat, it has at least one thousand hairs. Geach continues:

> Now let c be the largest continuous mass of feline tissue on the mat. Then for any of our 1,000 cat-hairs, say h_n, there is a proper part c_n of c which contains precisely all of c except the hair h_n; and every such part c_n differs in a describable way both from any other such part, say c_m, and from c as a whole. Moreover, fuzzy as the concept *cat* may be, it is clear that not only is c a cat, but also any part c_n is a cat: c_n would clearly be a cat were the hair h_n plucked out, and we cannot reasonably suppose that plucking out a hair *generates* a cat, so c_n must already have been a cat. So, contrary to our story, there was not just one cat called 'Tibbles' sitting on the mat; there were at least 1,001 sitting there! (p. 215)

Geach concedes, of course, that this conclusion is absurd, but it is interesting to observe wherein he professes to detect the fallacy. He explains:

> Everything falls into place if we realize that the number of cats on the mat is the number of *different cats* on the mat; and c_{13}, c_{279} and c are not three different cats, they are one and the same cat. Though none of these 1,001 lumps of feline tissue is the same lump of feline tissue as another, each *is* the same cat as any other: each of them, then, is a cat, but there is only one cat on the mat, and our original story stands. (p. 216)

Now I concede that this manoeuvre of Geach's saves the truth of the original story; but, as he says, there is a price to pay. 'The price to pay is that we must regard "- - - is the same cat as - - -" as expressing only a certain equivalence relation, not an absolute identity restricted to cats' (p. 216). Geach, however, is happy to pay this price, since he considers that it 'must be paid anyhow, for there is no such absolute identity as logicians have assumed' (ibid.), and in defence of this contention he refers us to earlier arguments in his book. I shall not consider those arguments here, though I do not as a matter of fact find them convincing. What I shall do, however, is to explain why I think that the truth of the original story can be saved far more plausibly without having to pay this price; and at the same time I shall try to show that Geach's resolution of the puzzle is in fact untenable.

Let me then say at once what my own solution to the paradox is. What I would say is that it is not merely *not* 'clear' (as Geach claims), but is in fact just *unintelligible* to suggest that the lump of feline tissue c (or any of the other lumps mentioned) *is a cat*. *None* of the 1,001 lumps of feline tissue is a cat, so there is not even a *prima facie* case for saying that there are 1,001 cats sitting on the mat. Why do I say this? Simply because the sortal terms 'lump of feline tissue' and 'cat' have *different criteria of identity* associated with them, and I contend that no individual of a sort ϕ can intelligibly be said also to belong to a sort ψ if ϕ and ψ have different criteria of identity. Geach clearly cannot believe this, or he would not have had to resort to his own means of dissolving the paradox, involving the abandonment of classical 'absolute' identity. (Incidentally, Geach himself explains admirably why sortal terms – or, as he calls them, substantival general terms – must have criteria of identity associated with them: see op. cit., pp. 63 f.)

But how can I defend my contention? More particularly, why do I say that no individual lump of feline tissue can *be a cat*? The answer is that I believe that one may construct a *reductio ad absurdum* of Geach's contrary assumption, as I shall now attempt to explain.

First of all, we should observe that Geach himself explicitly regards $c_1, c_2, c_3, \ldots, c_{1,000}$ and c itself as *different* lumps of feline tissue, so that he himself presumably subscribes to what we may call the *Lockean* criterion of identity for 'parcels of matter' or 'masses'. Locke, we may recall, asserts that 'whilst [a number of atoms] exist united together, the mass, consisting of the same atoms, must be the same mass, or the same body, let the parts be never so differently jumbled: but if one of these atoms be taken away, or one new one added, it is no longer the same mass, or the same body' (*An Essay Concerning Human Understanding*, Bk. II, Ch. XXVII, §3). According to this Lockean criterion (which also strikes me as being essentially correct), if x and y are parcels of matter of any sort (e.g. lumps of feline tissue), then x and y are the *same* parcel of matter if and only if x and y consist of the same atoms 'united together'. Thus c_n is a different parcel of matter from c simply because it does not contain the atoms included in the hair h_n, whereas c does.

Next we should notice that Geach also apparently subscribes to *something like* the Lockean criterion of identity for living organisms such as cats. Locke, it may be remembered, observes that '[i]n the state of living creatures, their identity depends not on a mass of the same particles; but on something else. For in them the variation of great parcels of matter alters not the identity' (ibid.). (Later, he goes on to suggest that this 'something else' upon which the identity of living organisms depends is 'a participation [in] the same continued life'; but precisely what this might mean need not concern us here, since all that does concern us is that Locke rightly ascribes *different* criteria of identity to parcels of matter on the one hand and living organisms on the other, and that Geach would apparently concede the rightness of this.) That Geach would subscribe to something like the Lockean criterion of identity for living organisms is apparent from that fact that, in defending his (in my view erroneous) contention that c_n is a cat, he says that 'c_n would clearly be a cat were the hair h_n plucked out, *and we cannot reasonably suppose that plucking out a hair generates a cat*' (op. cit., p. 215; my emphasis). The reason why this suggests that Geach subscribes to something like the Lockean criterion of identity for living organisms such as cats is that it indicates that he concurs with Locke (and common sense) in allowing that the gain or loss of a parcel of matter by a cat 'alters not the identity' of the cat (as Locke puts it).

So far, then, we have seen that Geach himself would appear to concede that the sortal terms 'lump of feline tissue' and 'cat' have *different* criteria of identity associated with them, and moreover criteria substantially akin to those Locke proposed for parcels of matter and living organisms respectively. We are now in a position to construct the *reductio ad absurdum* of Geach's supposition that the lump of feline tissue c (or any other such lump) *is a cat*.

The argument runs as follows. For the purposes of the *reductio* we are to suppose, with Geach, that the lump of feline tissue c is a cat. But c, being a lump of feline tissue, is a thing of such a sort that, according to the criterion of identity for things of that sort, it simply *ceases to exist* if various particles of matter (e.g. those contained in a certain hair h_n) are removed from it. (It ceases to exist because after the removal of these particles the atoms of which c was constituted are no longer 'united together'.) But nothing that *is a cat* can cease to exist merely on this account, because whatever is a cat must comply with the criterion of identity for living organisms, according to which (as Locke says) 'in them the variation of great parcels of matter alters not the identity'. Hence Geach is apparently committed to saying that c *both will and will not* cease to exist if various particles of matter are removed from it. So his assumption that c is both a lump of feline tissue and a cat commits him to *denying* that any particles of matter can be removed from c, and this is plainly false (and not merely contingently so). Therefore his original assumption must be erroneous.

We can generalize this argument to show that, *wherever* an individual x belongs to each of two different sorts ϕ and ψ, these sorts cannot have different criteria of identity (though they may of course have the *same* criterion and still be different sorts, as is the case with, e.g., the sorts *cat* and *mammal*). The point quite simply is that, if C_ϕ and C_ψ are the respective criteria of identity, we cannot in general rule out *a priori* the possibility that there should arise circumstances in which, according to C_ϕ (say), x will cease to exist, whereas according to C_ψ it will not.

All of this presupposes, of course, that if various particles of matter are removed from a lump of feline tissue such as c, then that particular lump of feline tissue *ceases to exist* (and, more generally, that the criterion of identity for a sort ϕ determines conditions under which individuals of that sort will cease to exist). But can this plausibly be denied? Certainly, after the particles of matter have been removed from *c*, there will be no *lump of feline tissue* with which *c* is any longer identifiable: of this much the Lockean criterion of identity for parcels of matter assures us. Geach, of course, would presumably want to say that *c* still exists *qua* cat, i.e. is still identifiable with the cat Tibbles, and so has *not* ceased to exist. But I find this suggestion quite unintelligible. *c* was originally individuated and introduced to us as a certain *lump of feline tissue*, and so I do not see how we can go on talking about *c* in the present tense once there is no such lump of feline tissue to refer to. (And, after all, if criteria of identity do *not* determine the conditions under which individuals of given sorts cease to exist, what on earth does? Or are we perhaps to relativize *existence* in the same way that Geach wants us to relativize *identity*? That seems to me the only resort available to him, and a particularly unpalatable one too. Geach should arguably have foreseen this, since it is a familiar fact that the concepts of identity and existence are intimately related logically.)

University of Durham © E. J. Lowe 1982

REPLY TO LOWE

By P. T. Geach

DR LOWE kindly sent his paper for comment; I am afraid I can only say that he commits an *ignoratio elenchi*. He ascribes to me a 'Lockean' view as to what criterion of identity we must use for a 'parcel of matter' over time; but he cites no passage from my writings parallel to his quotations from Locke, and indeed could not have cited one. It is integral to Locke's account to bring in atoms; an account that deserved the name 'Lockean' would at least have to bring in enduring fundamental particles of some sort; no theorizing that I am committed to does so. But in any case the identity over time of a lump or bit of feline tissue is clearly not given by any sort of 'Lockean' criterion for the diachronic identity of a 'parcel of matter'; in writing as if it were so given, Dr Lowe is oddly forgetting metabolism. Tibble's liver, or one of his teeth or hairs, remains the same bit of feline tissue during interchanges of matter; and the same goes for the bits of feline tissue mentioned in my argument.

University of Leeds © P. T. Geach 1982

REPLY TO GEACH

By E. J. LOWE

MY argument against Professor Geach does not, as he imagines, crucially rest upon my supposing that he assigns to lumps of feline tissue the Lockean criterion of identity for parcels of matter, so that I can accept with equanimity his denial of this. Rather it hinges on the surely undeniable fact that Geach is committed to assigning *different* criteria of identity to lumps of feline tissue on the one hand and to cats on the other.

Geach quite explicitly states that c and c_n (i.e. the lump of feline tissue which consists of c minus the hair h_n) are *different* lumps of feline tissue. Let us then suppose that the hair h_n is suddenly annihilated (a possibly which cannot be excluded by any thesis that Geach is here concerned to defend). In that case, I would claim, c (which was originally individuated and introduced to us simply as a certain *lump of feline tissue*) must cease to exist, since there no longer remains any *lump of feline tissue* with which c is now identifiable, i.e. which can be regarded as the *same* lump of feline tissue as c. It is true that c_n would still remain; but Geach has, as we have seen, already implicitly adopted a criterion of identity for lumps of feline tissue according to which c_n is *not* the same lump of feline tissue as c. (It really doesn't matter here whether this criterion is the Lockean one or not; I only surmised, in the absence of any explicit statement to the contrary, that Geach does adopt the Lockean criterion because that criterion certainly *would* deliver the verdict that c and c_n are different lumps of feline tissue.) But if c does not survive as c_n, it equally clearly does not survive as any of the other lumps of feline tissue still remaining after the annihilation of h_n. So it no longer exists.

Now, however, if Geach *also* wants to claim that c *is a cat*, he must allow that c *does* continue to exist after the annihilation of the hair h_n, since on any acceptable criterion of identity for cats the annihilation of a cat's hair cannot result in its ceasing to exist. That is to say, if c is a cat, then even after the annihilation of h_n there will still be some *cat* with which c is identifiable, so that c must still be in existence.

I conclude that Geach is committed to the absurdity of saying that c *both will and will not* continue to exist after the annihilation of h_n; an absurdity which can only be avoided by rejecting his assumption that c is *both* a lump of feline tissue *and* a cat.

University of Durham © E. J. LOWE 1982

REPLY TO LOWE'S REPLY

By P. T. GEACH

WE need different criteria to decide whether Tibbles, who exists at time t, still exists at time t', and whether a lump c of feline tissue, which coincides with Tibbles at time t, still exists at time t'; so much must be conceded to Dr Lowe. For my purposes c was supposed to be a *continuous* lump of feline tissue; the falling out of one single hair (we need not imagine it annihilated) ends the existence of c as a cohesive whole, but c's existence is different from the existence of Tibbles the cat. However, to allow this does not estop me from maintaining that at time t, being coincident with Tibbles, c *is* the same cat as Tibbles and thus *is* a cat. There would no doubt be inconsistency if I held that 'is the same cat as' is equivalent to 'is a cat that is (absolutely) identical with', or that 'is a cat' is equivalent to 'is identical with some cat'. But of course I have expressly rejected such equivalences. For me 'is the same cat as' just expresses a certain symmetrical transitive relation; and because this holds at time t between c and Tibbles, there is no logical reason why it should go on holding between c and Tibbles at time t'.

The matter may perhaps be clearer if we shift to my other example about men and heralds, since here proper names for heralds, unlike names for lumps of tissue, are in actual current use. Let us pretend that Hilary Handle is – *sc.* is the same herald as – Sable Basilisk. By 1984 Hilary Handel may have perished though there still is the herald Sable Basilisk, or contrariwise Sable Basilisk may have been "irreversibly" abolished by Mr Benn's Government although Hilary Handel still lives on. Obviously there is no inconsistency in such suppositions; nor is it necessary in order to describe such a history for us to say that, for some x, x perishes *qua* man but still exists *qua* herald, or perishes *qua* herald but still exists *qua* man. The situations here envisaged need no such out-of-the-way language for us to describe them.

University of Leeds

© P. T. GEACH 1982

CANONICAL COUNTERPART THEORY

By Graeme Forbes

IN a recent article in ANALYSIS ('The discreet charm of counterpart theory', 41.2, March 1981, pp. 73-6), Graeme Hunter and William Seager attempt to rescue counterpart theory (CT) from some objections of Allen Hazen's (see his 'Counterpart-Theoretic Semantics for Modal Logic', *Journal of Philosophy*, vol. 76, pp. 319-38). They see these objections as arising from 'uncritical use of the translation scheme originally proposed by Lewis', and intend to meet them by refraining from use of that scheme. But they do not offer a new scheme; they say '...it is no more necessary to have one to capture the sense of modal idiom than it is to capture the sense of quantificational idiom.... Appeal to truth value is the single most important criterion of correct translation' (op. cit., p. 72). Thus, where Lewis's scheme translates a truth by a falsehood or conversely, Hunter and Seager simply produce a sentence which they claim to be a better translation, without articulating any structural principles they employ to arrive at their candidate.

A friend of CT should not be happy with this procedure. Let L_m be first-order modal language and L_c be the language of counterpart theory. Sentences of L_c are intended to interpret or elucidate the meanings of L_m-sentences in the strong sense that they should be the outputs of an adequate theory of meaning, which we can formulate as a model theory, for L_m, written in L_c plus set theory. The interest of CT resides in its potential to provide an alternative approach to the model theory of modal logic, and from the clauses of any such alternative model theory we should be able to read off a translation scheme for L_m into L_c, just as we can in the standard case. If the motivation of CT is sound, which Hunter and Seager do not dispute, but nevertheless there is no translation scheme and so no counterpart-theoretic model theory, one might wonder if there is not some defect in the original first-order language L_m. And in fact, Hunter and Seager find certain L_m-sentences multiply ambiguous with respect to L_c; but if a counterpart-theoretic model theory is possible, to say that an L_m-sentence has more than one permissible L_c translation is just to say that we have not yet made up our minds what the model theory is to be.

To resolve these issues, we have to provide a model theory, or equivalently, a translation scheme into the language of a first-order theory of counterparts from which a model theory can be derived. However, the scheme must be free of the technical drawbacks of Lewis's, so as to be invulnerable to Hazen's objections. Such a scheme is provided below, prefaced with motivating remarks. In devising it, the modal logic S5 in the canonical formulation due to Kit Fine has been kept in mind (see Fine's 'Model Theory for Modal Logic Part I', *Journal of Philosophical Logic* (1978), pp. 125-56), and the main points about this formulation are these:

(i) there is no accessibility relation on the set W of worlds;
(ii) '□' expresses strong necessity, that is, the truth of '□A' at any world requires the truth of A at every world, not just at worlds in which the denotations of any terms in A all exist,
(iii) a predicate can be satisfied at a world w by objects which do not exist in w – thus, Jones is identical to himself even with respect to worlds in which he does not exist;
(iv) L_m contains a logically constant predicate 'Exists', whose extension at each world w in any model is the inner domain D_w of w;
(v) quantifiers are actualist, that is, a quantified variable ranges over only the inner domain of the world with respect to which the relevant formula is being evaluated.

L_c is, for convenience, a two-sorted language which contains the non-modal connectives of L_m, say negation and conjunction, and also the predicates and constants of L_m, each predicate except '=' boosted by one place. In deciding upon a counterpart predicate we take note of the distinction marked under (iii) above. Thus, counterparts of an object x must be available to satisfy certain predicates with respect to worlds at which x has no existing counterparts. One way of meeting this requirement is to stipulate that for each x in the set D of all possible objects in a model (the union of all the D_w), if x has no counterpart at w which is *in*, that is, which exists at, w, then x is its own counterpart at w; so the outer domain of every world w is D. (Whether this stipulation introduces trans-world identity is debatable, but if it does, such identities are in Kaplan's terminology conventional rather than representational features of the model.) From the way the stipulation is stated it can be seen that it is natural to have a *three*-place counterpart predicate 'Cv_1v_2w', which is to be read 'v_1 is a counterpart of v_2 at w'. So 'C(x, Jones, w)' means 'x is a counterpart of Jones at w', which leaves it open whether x is *in* w or not. The stipulation itself, which may be strengthened to a biconditional if 'and does not exist at w' is added to its last clause, is one of the axioms of CT, or alternatively, a clause of counterpart-theoretic model theory.

With these preliminaries, let us proceed to the translation scheme. We stipulate that the translation of any L_m-formula ϕ into a formula of L_c is the *relativization* of ϕ to the actual world w*. That is,

$$\forall\phi \in \mathrm{Form}(L_m), \mathrm{Trans}_{Lc}(\phi) = \mathrm{Rel}(\phi, w^*).$$

For each formula ϕ of L_m, we compute $\mathrm{Trans}_{Lc}(\phi)$ by means of the following recursive specification of $\mathrm{Rel}(\phi, w)$ for arbitrary w. (In the clauses which follow, 'Rel' is a functor which for particular values of the metalinguistic variables 'v', 'A', etc., yields terms for

L_c-formulae; corner quotes on both sides of the clauses are fulfilling their usual function of selective quotation; thus e.g. '$\ulcorner$-Rel(A, w)$\urcorner$' expands into '"-" $\frown$Rel(A, w)'.)

(1) If $F(t_1 \ldots t_n)$ is atomic, $\mathrm{Rel}(F(t_1 \ldots t_n), w) = F(t_1 \ldots t_n, w)$ except if F is '='; $t_1 \ldots t_n$ are either free variables or individual constants.

(2) $\mathrm{Rel}(\ulcorner {-A} \urcorner, w) = \ulcorner {-\mathrm{Rel}(A, w)} \urcorner$.

(3) $\mathrm{Rel}(\ulcorner A \,\&\, B \urcorner, w) = \ulcorner \mathrm{Rel}(A, w) \,\&\, \mathrm{Rel}(B, w) \urcorner$.

(4) $\mathrm{Rel}(\ulcorner \forall v A \urcorner, w) = \ulcorner (\forall v)(\mathrm{Exists}(v, w) \rightarrow \mathrm{Rel}(A, w)) \urcorner$.

(5) $\mathrm{Rel}(\ulcorner \exists v A \urcorner, w) = \ulcorner (\exists v)(\mathrm{Exists}(v, w) \,\&\, \mathrm{Rel}(A, w)) \urcorner$.

For each modal operator there are two cases, according to whether the formula it governs does or does not contain terms not within the scope of any other modal operator in that formula. If A contains no such terms, then

(6) $\mathrm{Rel}(\ulcorner \Diamond A \urcorner, w) = \ulcorner (\exists w_j)(\mathrm{Rel}(A, w_j)) \urcorner$.

But if $t_1 \ldots t_n$ are constants or free variables in $\Diamond A$ within the scope of the initial '$\Diamond$' but no other modal operator, then

(7) $$\begin{aligned} \mathrm{Rel}(\ulcorner \Diamond A \urcorner, w) = \ulcorner & (\exists w_j)(\exists v_1) \ldots (\exists v_n) \\ & (Cv_1 t_1 w_j \,\&\, \ldots \,\&\, Cv_n t_n w_j \,\& \\ & \mathrm{Rel}(A(t_i/v_i), w_j)) \urcorner \end{aligned}$$

where the v_i and world variable are peculiar to L_c and are chosen to avoid clash of variables.

There are two analogous clauses for '$\Box$':

(8) $\mathrm{Rel}(\ulcorner \Box A \urcorner, w) = \ulcorner (\forall w_j)(\mathrm{Rel}(A, w_j)) \urcorner$.

(9) $$\begin{aligned} \mathrm{Rel}(\ulcorner \Box A \urcorner, w) = \ulcorner & (\forall w_j)(\forall v_1) \ldots (\forall v_n) \\ & (Cv_1 t_1 w_j \,\&\, \ldots \,\&\, Cv_n t_n w_j \rightarrow \\ & \mathrm{Rel}(A(t_i/v_i), w_j)) \urcorner. \end{aligned}$$

Clauses (4) and (5) reflect point (v) about canonical S5, that quantifiers there are actualist; but it would be easy to accommodate possibilist quantifiers. Clauses (7) and (9) reflect point (ii), that '$\Box$' is strong, although it would be easy to accommodate a weak '$\Box$'. In addition, these clauses exhibit a sensitivity lacking in Lewis's original scheme, which is required for iterated operators. For example, in S5 we want '$\Diamond\Diamond Fa$' to be equivalent to '$\Diamond Fa$' even if the counterpart relation is not transitive. (6) and (7) give $\mathrm{Trans}(\Diamond\Diamond Fa)$ as

(10) $(\exists w)(\exists w')(\exists y)(Cyaw' \,\&\, Fyw')$

but the clause $T2_j$ of Lewis's scheme (simplified to a two-sorted language) gives $\mathrm{Trans}(\Diamond\Diamond Fa)$ as

(11) $$\begin{aligned} & (\exists w)(\exists x)(\mathrm{In}(x, w) \,\&\, Cxa \,\&\, (\exists w')(\exists y)(\mathrm{In}(y, w') \\ & \quad \&\, Cyx \,\&\, Fyw')). \end{aligned}$$

The L_c-cognate of (11) is not equivalent to (10), since it will speak of a counterpart of x at w', and this object, when the counterpart relation is not transitive, may not be a counterpart of a at w'. It is (10), of course, which is equivalent to Trans($\ulcorner \Diamond Fa \urcorner$).

What of the notorious L_m-sentence '$(\forall x)\Box(\exists y)(x = y)$', false relative to S5 with contingent existence but translated into a theorem of CT by Lewis's scheme? Our scheme translates it as

$$(12)\quad (\forall x)(\text{Exists}(x, w^*) \rightarrow (\forall w)(\forall z)(Czxw \rightarrow (\exists y)$$
$$(\text{Exists}(y, w) \,\&\, z = y)))$$

which is false for precisely the right reason. It is the introduction of the three-place counterpart relation and the inner/outer domain distinction which facilitates this result.

Next, what becomes of the inference from '$\Box Rab$' to '$\Box(\exists x)Rax$', where R is atomic? In their discussion of this inference, Hunter and Seager do not explicitly state the crucial point. We need to note that assuming actualist quantifiers, the inference is generally correct only if '$\Box$' is strong *and* the Falsehood Convention (that all non-logical atomic sentences are false at worlds where the things named in them do not exist) is in force. Our translation scheme for strong '$\Box$' affords the respective translations

$$(13)\quad (\forall w)(\forall z)(\forall t)(Czaw \,\&\, Ctbw \rightarrow Rztw)$$

and

$$(14)\quad (\forall w)(\forall z)(Czaw \rightarrow (\exists x)(\text{Exists}(x, w) \,\&\, Rzxw)).$$

(14) is not a consequence of (13) in first-order logic, but if we record in CT the fact that the Falsehood Convention is in force, perhaps by an axiom schema restricted to atomic substitutions, (14) will be a consequence of (13) in the resulting theory. And if we alter the translation scheme to make '$\Box$' weak, Trans($\Box(\exists x)(Rax)$) will not be a consequence of Trans($\Box Rab$) in any reasonable CT. These are the desired results. '$\Box Rab$' was a sentence which Hunter and Seager's methodology led them to pronounce quadruply ambiguous (op. cit., p. 74) but in fact their four readings correspond to four different sentences of L_m with strong '$\Box$', each of which has a single correct translation into L_c, as given by the scheme above.

Finally, what becomes of the necessity of identity? Although not discussed by Hunter and Seager, it was in difficulties with this formula that Hazen found the greatest inadequacy of Lewis's approach and the main motivation for his own complex proposals about the correct way of doing counterpart-theoretic semantics; and nothing we have said so far meets this part of Hazen's position. As is well known, the necessity of identity is translated into a theorem of CT only if each object has exactly one counterpart at each world, and this has to be added to the axioms of CT as a special stipulation, while the validity of the modal thesis itself is delivered by an intrinsic feature of standard S5 models, the feature that the transworld heirlines of a model are given by the real cross-world identities in it; from the usual point of view, if there are no such identities (with respect to inner domains) then every property of an object is essential to it. What this suggests, however, is that the advantage of counterpart theory may lie just in its suitability for contexts where transworld heirlines are not 'identity-like' and so cannot be given by any identity relation.

Hazen himself provides the clue to what such contexts might be like, when he points out (op. cit., p. 330) that a counterpart theory which simply permits branching heirlines has strange-seeming consequences if the usual 'Actually' operator is introduced into L_m (the reader will have no difficulty writing translation-scheme clauses for this operator). For instance, the following will be true:

$$(15)\quad \Diamond(\exists x)(A(\exists y)(x = y \;\&\; A(Fx) \;\&\; A(-Fx)))$$

although it seems to verge on contradiction. However on some views, there are contexts where the meaningful assertion that P is in respect of truth neither better nor worse off than the assertion that not-P, examples where we have a borderline case of P's obtaining. A thought prompted by this is that counterpart theory comes into its own when the possibility of being a borderline case of a counterpart of a given object arises, or better, where there is the possibility of *degrees* of counterparthood. If this is the only kind of case where the transworld heirline relation is not an equivalence relation which need be countenanced, the special case for which standard modal model theory is appropriate may be characterized as that in which it is always the case that a given object either definitely is or definitely is not a counterpart of another object. But there can be many other cases. Chisholm has pointed out that by making a sequence of small changes in an object which preserve identity we can accumulate a large change which does not preserve identity, which throws the coherence of transworld identity into doubt. But a counterpart relation which admits of degrees need not be transitive in any appropriate sense, and can be employed to construct a consistent model of the situation Chisholm has in mind (see his 'Identity Through Possible Worlds: Some Questions', *Nous* 1 (1967), pp. 1-8). In such a model, on an appropriate interpretation, (15) is of course not wholly true, but nor is it wholly false; rather, it has an intermediate, non-classical degree of truth related to the degrees to which the relevant counterpart relations hold. This is no more puzzling than the fact that on the envisaged approach to vagueness and borderline cases, employing essentially the Lukasiewicz infinite-valued logic, instances of 'P&-P' can take intermediate truth values; for in response to the question 'Is that person an adult?' asked of some precocious adolescent, we may perfectly coherently reply, 'Well, he is and he isn't'.

Tulane University, New Orleans © GRAEME FORBES 1982

WHAT WON'T ESCAPE SORITES ARGUMENTS

By PATRICK GRIM

PETER UNGER and Samuel C. Wheeler have recently used sorites arguments – adaptations of Eubulides' argument of the heap – in order to argue against the existence of swizzle sticks, stones, tables, people, and Peter Unger ([2], [3], [4], [5], [6]). Quine, too, takes sorites arguments of this sort seriously, and proposes that we abandon a wide range of apparently familiar objects in order to maintain a principle of bivalence ([1]).

Even those who have accepted these nihilistic conclusions, however, have thought that sorites arguments can be avoided fairly easily. Vague ordinary terms such as 'swizzle stick' and 'stone' fall victim to sorites arguments, and thus perhaps we must concede that there are no swizzle sticks or stones. But it has been assumed that we can easily introduce precise replacements for these terms which will escape sorites arguments. Unger, for example, anticipates a 'rather small' departure from common sense in the development of 'better, precise ideas' with which to replace the ordinary vague concepts which fall victim to sorites arguments ([4], p. 128). The only difficulty Unger discusses in this regard is that of deciding where to draw the line in a suitable replacement for, say, our ordinary term 'log';

> With respect to atomic removals ... at a given juncture in a given case, there are millions of removals which seem quite innocuous and favorable. The item resulting from one such does not seem any more 'loggy' than that resulting from any other of them. Which steps are to be ruled out, and why? ([4], pp. 149–50).

Quine proposes precise replacement as follows:

> When we do reach the point of positing numbers and plying their laws, then is the time to heed the contradictions and to work the requisite precision into the vague terms we learned by ostension. We arbitrarily stipulate, perhaps, how few grains a heap can contain and how compactly they must be placed. What had been observation terms are arbitrarily reconstructed, on pain of paradox, as theoretical terms whose application may depend in marginal cases on protracted tests and indirect inferences. The sorites paradox is one imperative reason for precision in science, among others ([1], p. 92).

In what follows I hope to show that sorites arguments, if taken seriously, are not to be avoided as easily as Quine and Unger suggest. The types of precise replacements which are usually envisaged would *not* escape sorites arguments. Any real escape, moreover, would demand a more complete and more drastic break from common talk of common things than those sympathetic to sorites arguments have generally anticipated.

I

Consider a typical sorites argument, taken from Unger:

> Here is an indirect argument to deny alleged swizzle sticks, those supposedly popular swizzle stirrers. We note that the existential supposition:
>
> (1) There is at least one swizzle stick,
>
> is inconsistent with the propositions we mean to express as follows:
>
> (2) If anything is a swizzle stick, then it consists of more than one atom, but of only a finite number.
> (3) If anything is a swizzle stick, then the net removal from it of one atom, or only a few, in a manner most innocuous and favorable, will not mean the difference as to whether or not there is a swizzle stick there.
>
> Supposing (1) and (2), by (3) we get down below two atoms and still say that a swizzle stick is there. That contradicts (2). The only way to maintain (2) and (3), while being consistent, is to deny the existence of those sticks ([3], p. 519).

So much for swizzle sticks, at least if we take the argument seriously enough to maintain bivalence and to accept (2) and (3) on grounds of linguistic intuitions. So much, too, for stones, tables, and people, for it is clear that similar arguments could be constructed against these.

Here precise replacements offer a promise of escape, however. Following Quine's and Unger's suggestions regarding limited-grain heaps and atomically-specified logs, we might propose precise '1-billion-atom swizzle sticks' as suitable replacements for vague 'swizzle sticks' *simpliciter*. '1-billion-atom swizzle sticks' would seem, at least at first glance, to escape sorites arguments. For in order to construct an argument against the existence of 1-billion-atom swizzle sticks analogous to Unger's argument above, it appears, we would need a premise (4):

> (4) If anything is a 1-billion-atom swizzle stick, then the net removal from it of one atom, or only a few, in a manner most innocuous and favorable, will not mean the difference as to whether or not there is a 1-billion-atom swizzle stick there.

But (4), unlike its predecessor (3), is obviously false; removal of even a single atom *will* mean the difference as to whether or not we have a 1-billion-atom swizzle stick. Without that crucial premise, it appears, sorites arguments will not succeed against 1-billion-atom swizzle sticks.

Sorites arguments are not to be avoided so easily, however. Consider a slight variant on Unger's argument, designed to show the non-existence of 1-billion-atom swizzle sticks. We first introduce the existential supposition:

(5) There is at least one 1-billion-atom swizzle stick,

and note that this is inconsistent with the following:

(6) If something is a 1-billion-atom swizzle stick, then something is a swizzle stick and is composed of 1 billion atoms.
(2) If anything is a swizzle stick, then it consists of more than one atom, but of only a finite number.
(3) If anything is a swizzle stick, then the net removal from it of one atom, or only a few, in a manner most innocuous and favorable, will not mean the difference as to whether or not there is a swizzle stick there.

Here (2) and (3) are the premises of Unger's original argument. (6) is an added premise which seems clearly true, since a 1-billion-atom swizzle stick would be a swizzle stick, composed of one billion atoms. But if we are to maintain consistency while accepting (6), (2), and (3), then we must deny (5), and with it the existence of 1-billion-atom swizzle sticks.

We might also argue for this conclusion somewhat less directly. Our original hope in introducing 1-billion-atom swizzle sticks as precise replacements was that (7) would be true:

(7) Sorites arguments fail to establish the non-existence of 1-billion-atom swizzle sticks.

But if this were true, together with the same apparently innocuous premise as before:

(6) If something is a 1-billion-atom swizzle stick, then something is a swizzle stick and is composed of 1 billion atoms,

we would have to conclude that:

(8) Sorites arguments fail to establish the non-existence of swizzle sticks.

If we follow Quine, Unger, and Wheeler in denying the existence of swizzle sticks on the basis of sorites arguments, then we will have to deny (8). Given the validity of the argument above, if we insist that (8) is false, we must deny (6) or (7). (6) is not a likely candidate. So we must deny (7) and concede that sorites arguments suceed in establishing the non-existence of 1-billion-atom swizzle sticks.

This conclusion should not be too surprising, despite the temptation to think, with Quine, that we can merely 'work the requisite precision into the vague terms that we learned by ostension' ([1], p. 92). For if there really are no heaps or swizzle sticks, there are no heaps of specific numbers of grains or swizzle sticks of precise numbers of atoms either.

II

1-billion-atom swizzle sticks are vulnerable to sorites arguments because they are still *swizzle sticks*, and are still specified using a term, 'swizzle stick', which falls victim to sorites arguments. If precise replacement is to offer an escape from sorites arguments, then, we at least shall have to be a bit more subtle in our introduction of precision. Let us thus attempt more careful specification of a precise replacement designed to escape sorites arguments. Without actually constructing the required definitions, we might outline the following strategy for defining 'gizzle gick', intended as a replacement for 'swizzle stick'.

Were we attempting to introduce a precise replacement for 'cubical object', we might begin with a mathematical specification for a perfect cube. In the present case we might start with a similarly ideal 'mathematically perfect swizzloid'. Imagine, if you will, a paradigm swizzle stick. A suitable definition for 'mathematically perfect swizzloid' would be one which specifies the abstract shape of that imagined paradigm in purely mathematical terms, much as a mathematical specification of a perfect cube specifies the abstract shape of an imagined cube.

'Gizzle gicks' could now be defined in terms of a specific range of physical approximations to mathematically perfect swizzloids of, say, five inches in length. We might first define 'earthly perfect 5″ plastic swizzloids' as non-empty sets of plastic molecules at a specific time, temperature, pressure, etc., which approximate the requirements of a mathematically perfect swizzloid as closely as is physically possible. Here further specifications would undoubtedly be required, but let us assume that with patience these could be included in our definition. 'Gizzle gicks' might then be defined so as to allow a specific range of variation. Something is a gizzle gick, we might propose, if and only if it either is an earthly perfect 5″ plastic swizzloid or is a non-empty set of plastic molecules at the specified time, temperature, pressure, etc., which would result in an earthly perfect 5″ plastic swizzloid by the addition of one hundred molecules or less, properly placed.

By some such elaborate means, it appears, we might be able to introduce an appropriate replacement for 'swizzle stick' without using in our definitions any term against which sorites arguments will prove effective. Plastic molecules would not themselves seem liable to gradual decomposition in the manner of sorites arguments, and non-empty sets of such molecules have a clearly specified lower limit; one molecule or more is necessary and sufficient for a non-empty set. If we take sorites arguments seriously, we will have to do without swizzle sticks. But gizzle gicks will stir our drinks (or ginks) equally well.

But will even this more elaborate introduction of precision offer an escape from sorites arguments? Perhaps not. For consider the following argument against the existence of gizzle gicks. We start with the existential supposition that

(9) There is at least one gizzle gick,

and note that this is inconsistent with:

(10) If anything is a gizzle gick, that is sufficient for it to be a swizzle stick, in the ordinary sense of the term.
(11) If anything is a swizzle stick, in the ordinary sense of the term, then it consists of more than one atom, but of only a finite number.
(12) If anything is a swizzle stick, in the ordinary sense of the term, then the net removal from it of one atom, or only a few, in a manner most innocuous and favorable, will not mean the difference as to whether or not there is a swizzle stick there, in the ordinary sense of the term.

Here (11) and (12) differ from Unger's original (2) and (3) only in the addition of 'in the ordinary sense of the term', which was clearly intended anyway. (10) is a new premise, but seems at least as compelling of assent as (11) and (12). If we construct the definitions suggested above properly, we will have defined 'mathematically perfect swizzloid' on the pattern of an imagined paradigm swizzle stick, and will have defined 'gizzle gicks' in terms of sets of plastic molecules approximating closely the requirements of a mathematically perfect 5″ swizzloid. So were I now to set a gizzle gick, suitably defined, before you, and were I to ask you whether that object would qualify as a swizzle stick in the ordinary sense of the term, I think you would have to answer 'yes'. Gizzle gicks have been specified in such a way that, were there such things, they would clearly qualify as swizzle sticks in the ordinary sense of the term. It *is* true that the plausibility of (10) relies on linguistic intuitions concerning 'swizzle stick'. But the same is true of (11) and (12), and of their predecessors (2) and (3). If we are to accept (2) and (3) in Unger's original argument on grounds of linguistic intuitions, there seems little reason not to extend (10) the same courtesy.

If we accept (10) through (12), however, we cannot consistently maintain (9) as well. Thus despite our care in constructing definitions with an eye to precise replacements, we must concede that there are no gizzle gicks.

We can also offer a more indirect argument for the same conclusion:

(13) Sorites arguments fail to establish the non-existence of gizzle gicks.
(10) If anything is a gizzle gick, that is sufficient for it to be a swizzle stick, in the ordinary sense of the term.
(14) Sorites arguments fail to establish the non-existence of swizzle sticks.

With Wheeler, Unger, and Quine, we have assumed (14) to be false. But (13) and (10) offer a valid argument for (14), so we must deny either (10) or (13). (10), for reasons given, seems a poor choice for denial. So it appears that we must deny (13). Despite whatever care we might take in constructing definitions such as those suggested above, it appears that precise replacements of this sort will not escape sorites arguments.

III

What might we offer as precise replacements which *would* escape sorites arguments? Here prospects seem significantly slimmer than proponents of sorites arguments have suggested. In order to escape arguments such as those presented above, any precise replacement would have to avoid analogues of (10). Thus nothing could escape sorites arguments the existence of which were sufficient for the existence of any of the common objects which seem to fall victim to sorites arguments. Nothing offered by way of precise replacement could, for example, qualify as a large object, a small object, a red object, or a non-red object, for each of these is subject to sorites arguments.

If we are still to hope for precise replacements which will escape sorites arguments, they will not be precise replacements of the simple sort which Quine and Unger suggest. What would be required would be not the mere addition of precision to existing terms, nor the 'rather small' changes in common thought which Unger envisages, but rather a fairly drastic departure from our common talk of common objects.

State University of New York at Stony Brook © PATRICK GRIM 1982

REFERENCES

[1] W. V. Quine, 'What Price Bivalence?', *Journal of Philosophy* LXXVII (1981), pp. 90-5.

[2] Peter Unger, 'I Do Not Exist'. In *Perception and Identity,* ed. G. F. MacDonald, Ithaca, New York: Cornell University Press, 1979, pp. 235-51.

[3] —— 'Skepticism and Nihilism', *Nous* 14 (1980), pp. 517-45.

[4] —— 'There Are No Ordinary Things', *Synthese* 41 (1979), pp. 117-54.

[5] —— 'Why There Are No People', *Midwest Studies in Philosophy* IV (1979), pp. 177-222.

[6] Samuel C. Wheeler, 'On That Which Is Not', *Synthese* 41 (1979), pp. 155-73.

PERSONAL IDENTITY AND PERSONAL SURVIVAL

By Andrew Brennan

'HOW to get a Jumbo from London to New York in 2 minutes', reads the heading over an advertisement. It turns out that only the plan of the aircraft and details about the seating arrangements are transmitted electronically in the two minutes. Disappointing, for these days we are not far from realizing the *Star Trek* fantasy of matter transportation. In this paper I look at the problem of whether we can survive, not journeys by a matter transporting beam, but 'journeys' involving no movement of our bodily matter at all. I start by looking at Parfit's view of survival, which seems to suggest that I could survive such a journey, and go on to argue that the notion of survival is no better off than the notion of personal identity, and to suggest a diagnosis for why Parfit thinks questions of survival may be more determinate than those of identity.

I The Parfit background

Parfit's central contribution to the personal identity debate can be enshrined in the following maxim: what matters for survival need not be identity. Consider the case (in Figure 1) where Anthea fissions into Beatrice and Cheryl:

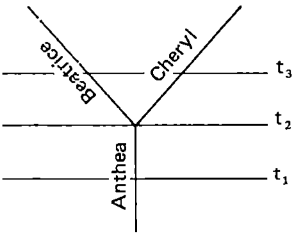


Figure 1

Just before t_2, the hemispheres of Anthea's brain are treated so they are almost identical in stored memory, linguistic disposition and habitual responses. At t_2 the hemispheres are separated, one staying in Anthea's skull, the other deposited successfully in an empty cranium attached to an otherwise healthy body. So as not to beg any questions, we christen the person who survives with Anthea's body 'Beatrice', the other 'Cheryl'.

Now clearly there are three answers to the question 'How is Anthea related to Beatrice and Cheryl?' They are: Anthea is Beatrice, Anthea is Cheryl, and Anthea is neither. (Each answer has its merits, and its difficulties.) Parfit, correctly in my view, says that there need not be a correct answer to the identity question in a case like this (1971a, p. 8). Although we think identity is germane to the question, 'Does Anthea survive after t_2?', Parfit suggests that the question of Anthea's survival can be settled prior to, and independently of, our decision on identity. Anthea survives, he maintains, both as Beatrice and as Cheryl, even though Beatrice and Cheryl are certainly not identical with each other.

The resurrection case discussed in Parfit 1971b gives us an extreme example of how to prise apart questions of identity and survival. Suppose God can – at the Great Day of Judgement – furnish replicas of people already dead. A replica will be a completely accurate reproduction of a person down to the finest biochemical detail. Unsurprisingly, the replica acts, talks, remembers and so forth in just the way the original did. Of course, there is no bodily continuity (nor continuity of brain material) in this case, and thus no associated causal explanation of why the replicas are psychologically continuous with their originals. (For an account of psychological continuity and connectedness, see Parfit 1971a, particularly pp. 11–21.) For those who see identity as the important thing, how can the existence of such replicas make sense of the notion of survival? Obviously, not at all. Change the focus, and see survival as the crucial thing, in no way presupposing identity, and this form of resurrection is as good as survival.

Parfit does qualify his view by indicating that survival is not, for him, an all or nothing matter. We can make sense of someone surviving to a greater or lesser degree as (or in?) someone else. If I wake up tomorrow with the ability to *q*-remember (Parfit 1971a, pp. 14–15) a large and significant part of the experience of Jane Austen, then Jane Austen, to some degree, can be said to survive as me. To a greater degree, presumably, Anthea survives both as Cheryl and Beatrice.

II FASTER THAN LIGHT JOURNEYS

A *Star Trek* style of transporter could have moved the Jumbo passengers, and perhaps even the jet itself, from London to New York in well under two minutes. Each passenger is broken down into his or her component elements, these are beamed at light speed across the North Atlantic, and the traveller is reconstituted at New York none the worse for wear. It seems clear that Parfit will admit that the traveller survives this process. And even those wedded to identity as the key notion can make perfectly good sense of what has happened. The bodily discontinuity of the traveller is just an extreme example of discontinuities bodies often suffer: a discontinuity of function alongside continuity of bodily stuff. A tent, a garden shed, a glider, and many other everyday objects are taken to pieces and put together again without us ever doubting their continuing self identity.

In a recent paper, Robert Elliot proposes a modified beam transport system. In his system, after the dissolution of the traveller's body into elements, these are stored in the departure terminal. All that passes from one terminal to another are blueprints enabling replicas of travellers to be constructed from the elements in store at the arrival terminal. Of course, those who emphasize bodily continuity will object that I do not – literally speaking – *travel* by the Elliot beam, since not one of my atoms leaves the departure terminal. For the Parfit-minded, however, the question is not whether I am identical with the replica but whether I survive as that replica. Not only is the replica psychologically continuous with me, but the cause of that continuity is perfectly evident. Indeed, the situation is not all that different from the resurrection case described in Parfit 1971b. By Parfit's account, I should – to some degree – survive the Elliot beam.

Elliot goes on, presumably ironically, to imagine arranging a journey between star systems that will be faster than light (Elliot 1981, pp. 4-5). He simply ensures that a copy of his blueprint is stored at one terminal together with instructions that he should be re-embodied at a specific time. Likewise, at the arrival terminal, a copy of his blueprint is kept, and precise instructions given for future embodiments there.

Suppose the unfortunate Anthea wishes to commute regularly between a small planet in the neighbourhood of Sirius (Sirius 4) and earth. Each interval of the time scale in Figure 2 represents the length of the journey at the speed of light between earth and

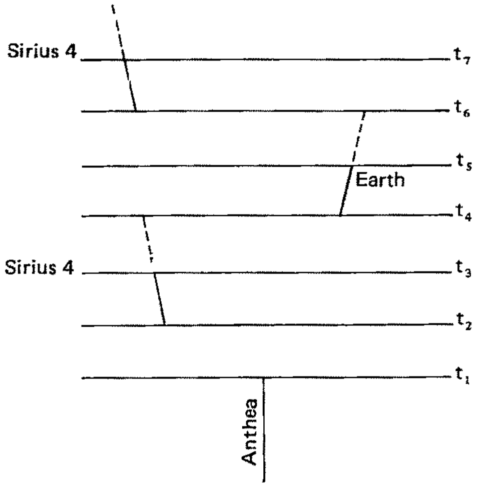


Figure 2

Sirius 4. On her first trip to Sirius 4, Anthea makes sure her blueprint is stored on earth, arranging to be re-embodied there just after t_4 (at t_4+5 seconds to be precise). Now she can stay on Sirius 4 until t_4-5 seconds (as shown by the dotted line), and so on for future journeys (provided she has her blueprint stored on Sirius 4 as well).

Unfortunately, Elliot's treatment of the case is flawed by his failure to recognize that the time saved in transit between the two terminals is bought at a cost. For the blueprint stored on earth only enables a replica of Anthea *as she was at* t_1 to be created at t_4+5 seconds. Nor is psychological continuity preserved in any of the journeys. So how on earth (so to speak) could Anthea survive the journeys between earth and Sirius 4?

III JOURNEYS AND 'JOURNEYS'

This narrow objection to Elliot perhaps misses the point. In the fission case we saw how Anthea was supposed to survive both as Beatrice and Cheryl. In our latest case, there is no time at which Anthea's replica on Sirius 4 competes with her replica on earth. Thus there are never two simultaneously realized candidates for identity with Anthea, and it was the presence of the rivals – Beatrice and Cheryl – which earlier persuaded us to prise apart the identity question from the question of survival. So without making an identity claim, might we not say that since Anthea survives both in her Sirius 4 and her earth embodiments she does, in a sense, 'journey' between the two terminals?

But we would only say this if we forgot to distinguish two separate claims. It is one thing to claim that Anthea survives both on earth and on Sirius 4; it is another to claim that she makes the journey from one to the other. Likewise, it is one thing to claim that the kind of resurrection already described is as good as survival; it is another to claim that something undergoes the process of resurrection, or survives between my death and the creation of my replica on the Great Day.

Sheds, tents, or gliders in their trailers, can survive journeys. All that happens is that the object undergoes some dismantling, and suffers some degree of functional discontinuity, while being shuttled from one place to another. In this way Anthea might undergo a *Star Trek* type of journey, beamed particle by particle to her destination. But, faced with the standard Elliot beam, the strongest claim that Parfit need make is that Anthea survives *as her replica*, not that she has undergone, or undertaken, the journey. Of course, we can always say that Anthea has survived the journey from earth, but we should have to put quotes round the term 'journey', or use that special tone of voice. What journeys from earth to Sirius 4 is not Anthea, but the blueprint: Anthea herself only 'journeys' from one to the other.

What applies to journeys by blueprint applies equally well to journeys at faster speeds. This is where Elliot has got things right. Whatever this metaphorical journeying is, it is the same, whether constrained by the speed of light or not. And here is the root of our puzzle. What notion of 'journey' are we using? It may help to consider the sorts of things that can go places – or, rather, can be *said* to go places – even though physical embodiments of them remain behind. Garden sheds, after all, only journey according to the movement of their physical stuff. But not so the design of a garden shed. The recipe for a Harvey Wallbanger only reached Aberdeen some time after it reached London. Oddly enough, as I hope to show in the next section, the travels of designs, recipes, ideas, beliefs, and so on are not irrelevant to our ponderings on personal survival.

IV TYPE AND TOKEN PERSONS?

We have already seen how Parfit prises the notion of identity away from that of survival. With equal dexterity, he prises memories away from the actual past experience of those who remember. This new notion of memory he calls '*q*-memory':

> I am *q*-remembering an experience if (1) I have a belief about a past experience which seems in itself like a memory belief, (2) someone did have such an experience, and (3) my belief is dependent upon this experience in the same way (whatever that is) in which a memory of an experience is dependent upon it (Parfit 1971a, p. 15).

Thus Beatrice and Cheryl *q*-remember Anthea's experiences. In the same way, my replica on the Great Day of Judgement will *q*-remember my past. Were *q*-remembering commonplace, as Parfit points out, our own memories would come to us as *q*-memories: 'Memories are simply *q*-memories of one's own experiences' (op. cit., p. 16). Better, then, to think of memories as having the double aspect of token and type. The same *q*-memory type can come to various people in the form of the appropriate *q*-memory tokens.

We think it distinctive of persons that they have the memories, intentions, ambitions and so on that they have. But the introduction of *q*-memories, *q*-ambitions, *q*-intentions and so on threatens the intimacy of this bond between a person and his or her 'inner' life. Just as the ideas expressed in a book transcend the copy of the book on your shelf, so Parfit seems to suggest that many of the properties of a person may transcend that person. Nor is this way of looking at our memories, intentions, ambitions and so on implausible. On the contrary, it captures one facet of our view of ourselves.

An author may look for a kind of immortality through his or her works. Many parents think they will not wholly die as long as their children survive, and politicians may seek to live on through the movements they found. Even if we count these notions as metaphorical, they are important enough not to be dismissed. They are important, also, for another reason. They give us the means to understand what Parfit says about survival.

Suppose, as in Dennett's 'Where am I?' fantasy (in Dennett 1979), the information processing structure of, and all the stored data in, my brain are fed into a computer by way of a massively complex program. The computer can now *q*-remember my past, and of course, all the details of my inner life can be displayed for the curious on a dozen VDU's. Were I to die tomorrow would I not to some degree survive, thanks to that computer program? And since the program is itself a type, not a token, and can be run on any of a dozen machines, it would seem that I might survive (in whatever sense) several times over. It is as though my death has been the death of one of my tokens, not of the person type, me. Certain types of cloning might likewise lead us to regard the members of a clone as tokens of the same person.

In both cases, there are grounds for arguing that the death of an individual is accompanied, in an important way, by the survival of something distinctive of (but not unique to) that individual. Parfit's resurrection by replication seems to be a case along the same lines. And the survival of Anthea through Cheryl's and Beatrice's *q*-memories, or through her replica on Sirius 4, seems. likewise to involve the same notion of 'survival'. With the possible exception of the resurrection case, Parfit's examples do not require the distinction between type and token persons, but they do require a distinction between type and token *q*-memories, and the rest. So, then, let us imagine that memories, intentions and ambitions can travel around in the way that beliefs, plans, recipes and so on can travel around. Do we have cause for comfort? Do we have good reason to believe that we survive the death of our bodies, or the Elliot beam?

The question is not an easy one to answer, and this for a very good reason. The notion of survival is vague enough for some, like Parfit, to think that the news is good, that the answer is 'Yes'. For others, the prospect of preservation on a computer program, of the survival of other members of one's clone, of the eventual emergence of a replica, give no cause for comfort at all. Are they misguided to be thus unconsoled? Of course not. The sensible policy here is to recognize that we are in the same impasse we faced when confronted with identity questions after the fission case. Parfit was happy to maintain that there was no question of right or wrong, no fact of the matter, as Quine would put it, to be settled in that case. What surprises me is that he should think there is a fact of the matter when it comes to the survival question. Let us agree that our notion of what a person is is quite indeterminate: no less is true of the notion of the survival of a self.[1]

University of Stirling

© ANDREW BRENNAN 1982

---

[1] I am grateful to my colleagues at Stirling for their criticisms of an earlier draft of this paper.

REFERENCES

Dennett, D. 1979. *Brainstorms* (Harvester Press).
Elliot, R. 1981. 'How to Travel Faster than Light', *Analysis* 41.1, January 1981, pp. 4–6.
Parfit, D. 1971a. 'Personal Identity', *Philosophical Review* 80, 1971, pp. 3–27.
Parfit, D. 1971b. 'On "The Importance of Self-Identity" ', *Journal of Philosophy* 68, 1971, pp. 683–90.

CAN WE TRAVEL FASTER THAN LIGHT?

By CHRISTOPHER RAY

ROBERT ELLIOT (ANALYSIS 41.1, January 1981) suggests that with judicious planning and by making use of advanced technology anyone can travel faster than light. The technological requirements are the abilities (i) to manufacture a 'total' biochemical blueprint of the physiological and psychological composition of a given individual, (ii) to reconstruct the individual from the details of the blueprint even after the transmission of these details to distant star systems. Elliot tells us how we can make the appropriate plans for faster than light travel between points four light years apart. We shall look at the experience of Mr E who wishes to travel this distance, from X to Y, using the Elliot method.

(1) At point Y a total blueprint of E is made and stored;
(2) E has given instructions that he should be embodied from the details of the blueprint after a precisely given time – say exactly ten years;
(3) E then travels by 'conventional' means to point X – according to Elliot the details of the blueprint are transmitted at the speed of light from Y to X, E being disintegrated at Y and embodied at X; this transmission takes four years;
(4) after spending six years in the vicinity of X, E is then disintegrated at X;
(5) at the same time, in accordance with the instructions given at (2) above, E is embodied from the details of the blueprint at Y;

(6) the total transit time for E's 'journey' from X to Y is zero; hence E has travelled faster than light – at the speed of light the transit time would have been four years as at (3) above.

Elliot assures us that the E arriving at Y (E_y) is 'psychologically continuous and cohesive' with the E who 'left' X (E_x) (p. 5). This is the case despite the lack of causal continuity of E between X and Y: causal signals, we shall agree, can travel no faster than light. Presumably Elliot also believes that the 'two' E's are physiologically continuous and cohesive. Thus the E's are mentally and physically one individual. The journey from X to Y is merely an instantaneous change of spatial position for this individual.

I don't wish to quibble with the assumption that the simultaneity of distant events is well-defined in the relativistic context which Elliot tacitly adopts. Nor shall I challenge his presumption that it is possible in principle to make a physical blueprint of all aspects of an individual, both physiological and psychological. I shall concern myself only with the claim that E_x is physically and mentally continuous and cohesive with E_y. If this is not the case, we then have no reason to suppose that E_x travels to Y.

Stage (4) above is crucial. We would expect that E_x would undergo many physiological and psychological changes during these six years. He might fall in love or develop a fear of spiders; he might lose a tooth or acquire an ulcer. Consequently the physiological and psychological composition of E_x after six years at X will differ in many respects from the composition of E_y as recorded on the blueprint at Y. Consider the extreme case in which E_x dies just before the moment designated for his disintegration at X: the very lively E_y embodied at Y could hardly be said to be continuous and cohesive with the corpse at X. But even in marginal cases in which E_x spends seconds or fractions of a second rather than years at X before his disintegration there and the simultaneous embodiment of E_y at Y, changes will occur: perhaps E_x will acquire a memory of his short stay; there will certainly be a number of physiological changes, e.g. the death of cells or the firing of neurons in the brain. Even if we accept that E_x at the moment of his arrival at X is the same in all respects except spatio-temporal position as E_y at the moment of his embodiment at Y, as time passes for E_x the differences will grow between him and his alter ego. In short, if E_x spends any time at all at X before his disintegration there, then E_x cannot be the same individual as E_y. This means that we lose the psycho-physiological continuity essential to Elliot's claim that in such circumstances E can travel faster than light.

We could only have continuity between E_x and E_y if no time elapses from E_x's point of view between his arrival at X and his disintegration at X. But this would only be the case if E_x is not embodied at X at all. Then there would be no E_x for us to disintegrate. All that has happened is that: (a) E_y has been disintegrated at Y and embodied at Y from the details of the blueprint stored at Y in accordance with the instructions given at Y, and (b) a copy of the blueprint details has been sent from Y to X. There is no question of anyone travelling from X to Y. For there is no E_x available for any journey. And as soon as we make E_x available he begins to undergo changes which break the psycho-physiological continuity between him and E_y.

Balliol College, Oxford

© CHRISTOPHER RAY 1982

A NOTE ON THE SYNONYMY PRINCIPLE OF PROPERTY IDENTITY

By MICHAEL TYE

MANY philosophers dismiss the traditional view that properties are identical if, and only if, they are expressible by synonymous predicates either on the grounds that it is inconsistent with such everyday contingent property identities as 'Red is the colour of ripe tomatoes' or because of an alleged conflict with certain scientific reductions, for example, 'Temperature is mean molecular kinetic energy'.[1] I have argued elsewhere that this response is too hasty: a proper understanding of the traditional view shows that it can accommodate both types of case.[2] In this note, I wish to examine three further objections to the traditional position.

The principle of property identity which I shall defend (with some qualifications to be introduced shortly) may be given a more precise formulation as follows:

> (I) Property P = property Q iff given any predicates F and G if P is expressible by F and Q is expressible by G then F is synonymous with G.

One objection to (I) is that it is circular. Predicates are linguistic *types*, since one and the same predicate may be used by different speakers. Hence predicates are really properties, that is, entities which may be attributed to the objects of reference (as, for example, in 'That word token is an instance of "is red"'). It appears, then, that the right hand side of (I) quantities over properties; and this means that (I) is circular.

---

[1] For the former criticism, see, e.g. Jaegwon Kim, 'On the Psycho-Physical Identity Theory', *American Philosophical Quarterly* 3 (1966), pp. 232–33; also 'Events as Property Exemplifications', in *Action Theory*, ed. by M. Brand and D. Walton (Reidel, Dordrecht, 1976), p. 168. For the latter criticism, see, e.g. Max Deutscher, 'Mental and Physical Properties', in *The Identity Theory of Mind*, ed. by C. F. Presley (University of Queensland Press, 1967), pp. 74–5; Hilary Putnam 'On Properties', in *Essays in Honor of Carl G. Hempel*, ed. by Nicholas Rescher (Reidel, Dordrecht, 1969), pp. 235–7.

[2] See my 'On An Objection to the Synonymy Principle of Property Identity', *Analysis* 41.1 (January 1981), pp. 22–6. Also my 'Scientific Reduction and the Synonymy Principle of Property Identity', *Philosophical Studies*, 40.2 (1981), pp. 177–86.

Some may find this kind of circularity untroubling. After all there are other principles of identity which are similarly circular. Take, for example, Davidson's criterion of event identity: events are identical if, and only if, they have the same causes and effects. Still in view of the fact that many philosophers *require* adequate identity conditions to be noncircular,[3] we would do well to consider ways in which we can avoid the charge of circularity.

The simplest way to meet the objection is to revise (I) so that it becomes

> (I′) Property P = property Q iff given any predicate tokens F and G if F expresses P and G expresses Q then F is synonymous with G.

In (I′) talk of predicates has been replaced by talk of predicate tokens. Since predicate tokens are particular, concrete, dateable entities they cannot be classified as properties. Hence (I), in its revised form, is clearly not circular.

This response to the first objection, natural though it is, brings to the fore another objection which seems to run deeper. Properties constitute a nondenumerable infinity, since the real numbers are indenumerably many and for each real number n there are such properties as being greater than n or being less than n. Predicate tokens, however, form a finite class. Hence, there are nonidentical properties which are never expressed in predicate tokens. But, according to (I′), all linguistically unexpressed properties are identical.

This objection to (I′) cannot be answered merely by shifting back to (I) and swallowing its circularity. For the totality of predicates in any humanly possible language can be no larger than a denumerable infinity, that is, one the size of the totality of natural numbers. Hence, the objection to (I′) is applicable *mutatis mutandis* to (I).

It seems to me that this objection is decisive if the 'if-then' connective on the right hand sides of (I) and (I′) is taken to be material implication. But suppose we view this connective as having counterfactual force. Then the preferred principle (I′) can be rewritten as

> (I″) Property P = property Q iff if there had been a predicate token F and a predicate token G such that F expressed P and G expressed Q than F would have been synonymous with G.

---

[3] See, e.g. Neil L. Wilson, 'Facts, Events and their Identity Conditions', *Philosophical Studies* 25 (1974), pp. 303-21; George Sher, 'On Event Identity', *Australasian Journal of Philosophy* 52 (1974), pp. 39-47.

Understood in this way, (I′) does not entail that all linguistically unexpressed properties are identical. Nor does it require that there be less than a nondenumerable infinity of properties. For with respect to any two different properties which involve relations to real numbers, it will not be true that if there had been predicate tokens which expressed those relational properties then the tokens would have been synonymous.

Some philosophers may feel that the introduction of counterfactuals is too high a price to pay for the defence of the traditional position. I do not agree. For one thing, counterfactuals are not as opaque to us as they once were. Plausible truth conditions are now readily available.[4] For another how much we're prepared to pay should depend upon what the market has to offer. (I″), it must be remembered, provides us with a *completely general* set of necessary and sufficient conditions for the identity of properties. Other principles of property identity may be simpler but frequently they purchase this simplicity by restricting themselves to just one kind of property (for example, the sort found in scientific theories[5]). If (I″) is deemed too complex (or vague) then the onus is upon the philosopher who rejects it to come up with a more straightforward principle which operates at the same level of generality and which is not open to obvious counter-examples.

The final objection I want to consider consists of a counterexample to (I) which stands apart from most of the stock cases in the literature. Let both *P* and *Q* be the property of killing John, let *F* be the predicate 'kills John' and *G* the predicate 'kills himself'. Obviously *F* and *G* are not synonymous, yet *P*, *Q* is expressible by both, as in 'John kills John' and 'John kills himself'.[6]

This objection, which may be extended to the revised versions of (I), can be handled in two ways. To begin with, we might reasonably challenge whether the property expressed by 'kills himself' in 'John kills himself' is really the same as the property expressed by 'kills John' in 'John kills John'. The former property, we might argue, can be correctly attributed to persons who do not kill John, as, for example, in the true sentence 'George kills himself' while the latter cannot. Hence, by Leibniz' Law the properties are not identical. It seems to me that this line gains a good deal of plausibility from the thought that if George and John kill themselves, then there is some property which they *share*, some property other than merely killing something. What is this property? It can't be killing George nor can it be killing John. Instead the common property here is surely the property of self killing. This is the property which tokens of 'kills himself' express, and *this* is what George and John share.

---

[4] See, e.g. David Lewis, *Counterfactuals* (Harvard University Press, Cambridge, Mass., 1973).

[5] For an example of such a principle, see the one defended by Hilary Putnam in 'On Properties', op. cit.

[6] I owe this counterexample to Herbert Heidelberger.

Another possible response for those who are not persuaded by this initial reply is to restrict the predicates or predicate tokens spoken of in our three versions of the synonymy principle to those which avoid personal pronouns and indeed demonstratives generally (be they temporal, spatial, experiential, or personal). Taking this stance, we can say on behalf of the final principle (I″), for example, that the property expressed by the predicate token in the utterance 'John kills John' *is* the same as the property expressed by the predicate token in the utterance 'John kills himself', since the only real property here is killing John and this property is such that if there had been different entirely *nondemonstrative* predicate tokens which expressed it, then they would have been synonymous.

I conclude that the traditional position on property identity has more merit than is generally realized. The revised principle, (I″), is, I suggest, the most plausible unrestricted account of property identity we have.[7]

Northern Illinois University © MICHAEL TYE 1982

---

[7] I would like to thank John Barker for a helpful suggestion.

UTILITARIANISM AND RATIONALITY

By EARL CONEE

IN 'A Utilitarian Paradox' (ANALYSIS 44.2, March 1981, pp. 105–12) Frederick Kroon offers an example with which he challenges the coherence of utilitarianism. I shall argue here that the incoherence is better understood to lie elsewhere. The essential components of Kroon's example are these assumptions:

> A rational utilitarian, S, is on the point of judging the moral status of an act, A, which S knows to be a case of gratuitously inflicting pain. S also knows that the mental occurrence of his judging A to be wrong would have tremendously good consequences which far outweigh the evil of A. And S knows that his failing to evaluate A as wrong – either by judging A to be right or by making no moral appraisal of A – would have bad effects.

Concerning this case Kroon argues as follows. Judging A to be wrong seems at first to be what S ought to do. But S knows that the beneficial effects of his rendering that judgment would count among the consequences of A, making A morally right according to S's utilitarianism. 'A negative evaluation is known by [S] to be incorrect, therefore, and so [S] cannot judge that A is wrong.... Nor can [S judge that A is right,] since that evaluation is also known by [S] to be incorrect.' (Recall that S knows the bad effects of his not judging A to be wrong, consequences which would merely render an evil act so much the worse.) Yet S '... can scarcely affect moral indifference or refuse to commit [himself] on the question of A's morality, [for] in that case A will be known by [S] to be wrong.' Kroon summarizes the problem in this way: '[S] is unable to proceed, for [he] has a string of counterfactuals whose structure precludes the making of *any* evaluation: were [S] to evaluate A as (1) wrong, (2) right, (3) indifferent, A *would* be (1) right, (2) wrong, (3) wrong.'

Kroon canvasses a few ways to 'draw the sting' from the problem, and finds none of them successful. He suggests that '... some standard utilitarian ideas need to be revised or discarded if utilitarianism is to stand as a minimally coherent theory'.

Let us bring into sharper focus the assumptions relied upon to argue for the incoherence. The reason given for S's inability to judge that A is wrong is that S knows that evaluation to be incorrect. More precisely, what S knows is this: if S judges that A is wrong, then, because the enormously good effects of S's making that judgment redound to the credit of A, A is right. Presumably we are to see that such knowledge would render a rational individual incapable of forming the judgment that A is wrong. An important assumption, then, is that S has a sufficiently rational constitution to satisfy something like:

> P1 While S knows that S's judging x to be W implies that x is R, and knows that being W is incompatible with being R, S is unable to judge that x is W.

Another similar assumption is required. Using P1 and the knowledge given S we can infer that S is unable to judge A to be wrong, and, by parallel reasoning, that S is unable to judge A to be right. Kroon holds that S also cannot refrain from making one of those judgments. He infers this from the stipulation that S knows that his not making either evaluation is a sufficient condition for A's being wrong. So the argument depends upon something like the following also being true of S:

> P2 While S knows that S's failing to judge x to be W and failing to judge x to be R implies that x is W, S is unable to refrain from making one of those judgments.

We now have before us all the makings of an impossible situation. Rejecting utilitarianism appears to provide a solution to the paradox, since then S would not know that his mental state affects the moral status of A as utilitarianism implies. But that would not solve the general problem that this sort of case raises. P1 and P2 have been formulated so as to apply to judgments other than utilitarian evaluations of acts. Here is another case where they make trouble, this one having nothing to do with morality:

> A thirtieth century brain physiologist, T, knows that all of a person's N-fibres fire if, and only if, the person judges that not all his N-fibres ever fire when he considers the matter for the first time. Knowing this, T considers for what he knows to be the first time: (a). All of T's N-fibres will fire.

By replacing 'S', 'x', 'W', and 'R' in P1 and P2 with 'T' '(a)', 'false', and 'true' respectively, we can see that T is unable to judge that (a) is false. And when we switch 'false' and 'true' in P1, we see that T is unable to judge that A is true. The new P2 tells us that T is also unable to refrain from judging the truth-value of (a), since his physiology assures him that failure then to evaluate (a) as true or as false renders it false. Paradox.

Such reasoning can be used against any theory that allows the following sort of case: Knowledge of the theory and of what the consequences of certain mental acts would be gives knowledge that evaluating something implies that evaluation to be incorrect, and that failing to evaluate the thing implies a particular evaluation to be correct. Basically the same problem would remain even if all such theories were counted false. Knowledge is not actually crucial. Justified belief in the theory seems as plausibly to be capable of rendering a rational subject unable to form, or withhold, the relevant judgments. Thus in order to block the paradoxes by faulting the theories we would have to go to the extreme of claiming that they are all insufficiently coherent to be the basis for rational belief in the problematic implications.

A much better solution is available. We justified attributing P1 to S by assuming him to be bound to adhere to rational scruples in adopting beliefs. But that assumption does not warrant attributing P2 to him. Knowing that failure to evaluate A implies that A is wrong does not by itself justify any particular evaluation of A. Such knowledge is compatible with, e.g., reasonable doubt whether A is right. If S is ignorant of whether or not he *does* fail to evaluate A during the interval when such failure is known to suffice for A's being wrong, that known sufficiency creates no rational requirement for S to evaluate A. What S knows, together with his staunch rationality, seems to compel him to judge that A is wrong only when he also knows himself to withhold evaluation while doing so ensures that A is wrong. Then he knows the premises of a *modus ponens* argument to the conclusion that A is wrong. So we should add to the case that S knows that he does not evaluate A while that affects A's moral status, and investigate further.

Three types of cases call for separate consideration. S might have the knowledge in question *during* the relevant interval. We can bring that about by assuming that S foresees that his ironclad disposition against knowingly rendering a false judgment will cause him to refrain from evaluating A throughout the pivotal interval. But in doing this we have given the case extraordinary characteristics which assure us that S does manage to refrain for a time from drawing a direct inference from known and considered premises. To make P1 true of S we had to make him one who does not evaluate A during the crucial interval. Paradox threatens if S's rationality also seems to show that he evaluates A. However, we now see that to supply the knowledge which establishes for S that a particular evaluation of A is correct we must have him know that he cannot form a belief which he sees would be incorrect. Thus we have seen to it that he cannot do so even when *modus ponens* reasoning provides the inducement.

This is not to argue that suspending judgment when one's evidence powerfully favours a given belief is more rational, or easier, than adopting a belief against convincing evidence to the contrary. Rather, in the case at hand S has all the evidence for the self-falsifying effect of evaluating A whether he judges or suspends judgment. In contrast, S's knowledge that suspension renders A wrong must be supplemented by knowledge that he does suspend judgment for S's evidence to prove that A is wrong. Thus when we fill out the case so as to make the evidence for believing A to be wrong as good as the evidence that so judging would be self-falsifying, we must make it known to S that he fails to evaluate A. So to get as close as we can to making S subject to conflicting dictates of reason we are required to have him know himself to have what it takes not to evaluate A when that would be a mistake. And making this so has the result that S's rational compunction against knowingly judging falsely wins out over the rational incentive to believe direct implications of what he knows. Since we are thus assured that his rationality does not compel belief despite prima facie compelling evidence, and nothing else about the case implies that S evaluates A, we have no paradox in this type of case.

Suppose S has mere justified belief that he suspends judgment when that matters — S's knowledge of his rationality warrants the belief that he will not then evaluate A, so both premises of the *modus ponens* argument are justified for him; but he incautiously proceeds right away to draw the conclusion that A is wrong. In that case we have falsified the attribution of Pl to S. For then S does render a judgment which he knows is false if he so judges. Perhaps his inclination to accept clear implications of what is evident to him leads him to fail to attend to the effects of making the inference. Drawing such a conclusion would be an instance of a routine sort of fallibility in an extremely peculiar evidential situation. It would not be a paradoxically irrational lapse. The coherence of utilitarianism is not jeopardized by this sort of case, either.

The remaining cases are those in which S does not realize until the *end* of the pivotal interval that he suspends judgment throughout. At that point, his promptly drawing the conclusion that A is wrong would be unproblematic, for it would not have the effect of making A right. What about the very last moment in the interval (assuming that it has one, and that S knows it to be such)? S might be someone who takes no time to accept immediate consequences of what he comes to know – it would be implausible to insist that no one could possibly do that. So why is it not paradoxical that S must take time to draw the conclusion is this case? The answer is that the acknowledged possibility of simultaneous acceptance of propositions and their obvious implications does not extend to implications derived from what one does not accept. When a known lack of belief in a statement helps to establish the truth of the statement for a person, the person must change in order to come to believe it. And change must take time. So it is no wonder that S believes that A is wrong only after the interval of suspending judgment that helps to establish for him that A is wrong.

However we develop the details of the example, there is no way to make it plausible that both P1 and P2 apply at once. S's being rational never requires their joint satisfaction. Thus such an example does not justify our concluding that utilitarianism (or any of the other susceptible theories) is defective.

Indiana University © Earl Conee 1982

THE ARGUMENT FOR ANOMALOUS MONISM

By Ted Honderich

DONALD DAVIDSON's Anomalous Monism, his engrossing Identity Theory of the mind,[1] emerges from reflection on what seems to be a contradiction. The seeming contradiction is a matter of three claims, the first of which is that there are causal connections between physical and mental events. The second is The Principle of the Nomological Character of Causality. Wherever there are causal connections between events, the events are connected by law. The third claim is that there are no psychophysical lawlike connections. We escape the seeming contradiction and get to Anomalous Monism by way of a certain understanding of the second claim..

---

[1] 'Mental Events', in Lawrence Foster and J. W. Swanson, eds., *Experience and Theory* (London: Duckworth, 1970). See also 'Psychology as Philosophy', in S. C. Brown, ed., *Philosophy of Psychology* (London: Macmillan, 1974) and 'The Material Mind', in P. Suppes, L. Henkin, G. C. Moisil, A. Joja, eds., *Philosophy of Science IV* (Amsterdam: North-Holland, 1973). The three papers are reprinted in Donald Davidson, *Actions and Events* (Oxford: Clarendon, 1980), from which I take page references to all of Davidson's papers mentioned. I am most grateful to Davidson for comments on an earlier draft of this essay, which led me to enlarge it. He is not responsible for the upshot, and it does not have his agreement. My thanks too to Colin McGinn.

While causal connection holds between events however described, 'laws are linguistic; and so events can instantiate laws, and hence be explained or predicted in the light of laws, only as those events are described in one way or another. ... the principle of the nomological character of causality must be read carefully: it says that when events are related as cause and effect, they have descriptions that instantiate a law. It does not say that every true singular statement of causality instantiates a law'.[2] When E_1 and E_2 are cause and effect, it does not follow that they are in lawlike connection as E_1 and E_2, or under the descriptions 'E_1' and 'E_2'. The two events may be so connected but need not be.

If a mental event causes a physical event, they can therefore be in lawlike connection under other descriptions. Given the third claim, that there are no psychophysical lawlike connections, any such mental event *must* be in lawlike connection under some other description. It must be so as a physical event. We therefore have an identity. However, the third claim comes into operation a second time at this point, and so what we have is a lawless identity. It is not a matter of law that the mental event is what it is, identical with a physical event. As the idea of Anomalous Monism is also expressed, on the assumption that types bring in law, we have token-identity but not type-identity.

Davidson has elaborated his conception of an event as an irreducible entity, not something to be removed from our ontology. An event, further, has an indefinite number of properties, features or aspects. When I speak of Dora's fall, I do not in those words fully describe all of that event. It was a fall from the top of the table, during her birthday speech, and so on.[3] This conception of events is not necessary to the proposition that causal connection as ordinarily conceived need not entail a lawlike connection in the same terms, but it will be simplest to proceed by way of the given ontology. Certain it *is* true that when I put some pears on the scale, something green and French did cause the pointer to move to the two-pound mark, but there in fact is no entailed law connecting greenness and Frenchness with the pointer's so moving. There is in fact no law at all connecting the event in virtue of its being of something green and French with the pointer's moving to the two-pound mark. There is no lawlike connection connecting the first event in virtue of greenness and Frenchness with the second event in virtue of its being the pointer's moving to the two-pound mark.

It is to be noticed that we have given clear sense to talk of something's being such and such *as something or other*, or *under a description*. To talk this way is to speak of certain properties of a thing rather than others. To say two things are not in lawlike connection under certain descriptions is to say that certain of their properties are not in lawlike connection, or, perhaps, that the things are not in lawlike connection in virtue of certain of their properties. Perhaps everyone has always understood 'under a description', and 'as' when it is so used, in this way. So far as I can see, Davidson does not disagree.[4] Certainly that lawlike connection holds in virtue of certain properties is not in conflict with the line that 'laws are linguistic', understood as it must be. Davidson's doctrine of the nature of laws[5] is not fully developed, but presumably it takes lawlike connection to be no more a matter of language than causal connection is made so by the line that 'causal statements are linguistic'.

---

[2] 'Mental Events', *Actions and Events*, p. 215.

[3] 'The Logical Form of Action Sentences', in Nicholas Rescher, ed., *The Logic of Decision and Action* (Pittsburgh: Pittsburgh University Press, 1967); 'Events as Particulars', *Nous* 5, 1971.

To return to the event of the pears, there is no denying that it is only certain properties of the event which are relevant to its being the cause it is. Davidson asserts, certainly, as already noted, that the substitution of coextensive descriptions in causal statements does not affect truth value.[6] We must distinguish between a cause and the feature we hit upon for describing it.[7] All of that, however, is consistent with the truth that neither the greenness nor the Frenchness of the pears was relevant to the event of the pears' being put on the scale in so far as that event caused the pointer to move to the two-pound mark. So with effect-events. That the scale's pointer was made in Sheffield is irrelevant to the event of the pointer's moving to the two-pound mark being the effect in question. John Mackie sets out the fact of causally-relevant properties clearly.[8]

Is there a difficulty in the idea that it is in virtue of certain of its properties rather than others that an event is the cause it is? Well, the event of the pears' being put on the scale would not have been the event it was if the pears were not green and French. Thus there would be a barrier to saying that *that* event would have caused the pointer's moving to the two-pound mark if the pears had not been green or French. *That* event would not have occurred. Does that make the pears being green and French causally relevant to the given effect?

It seems to me clear that it does not. Certain conditional connections hold between the weight of the pears and the pointer's movement, and they do not hold between the greenness or Frenchness of the pears and the pointer's movement.[9] The greenness and Frenchness were necessary to the event's being the event it was, but not necessary to the event's being the cause it was. Certainly it may be said that the cause that there was would not have existed if the pears have not been green and French. That is consistent with the greenness and Frenchness being causally irrelevant to the effect. That we say, as we do, that the cause that there was would not have existed if the pears had not been green and French is owed to a fact of language – roughly the fact that we take the whole for the part – and not to any fact of causal necessity about all properties of the pears. There is no such fact.

---

[4] 'Eternal vs. Ephemeral Events', *Nous* 5, 1981. *Actions and Events*, pp. 194–5.

[5] 'Emeroses by Other Names', *Journal of Philosophy* LXIII (1966).

[6] 'Causal Relations', *Journal of Philosophy* LXIC (1967). *Actions and Events*, p. 152.

[7] 'Causal Relations', op. cit., pp. 155–6; 'Mental Events', op. cit., p. 215.

[8] *The Cement of the Universe* (Oxford: Clarendon, 1974), p. 260. See also Ch. 3.

[9] An analysis of causation in terms of certain conditionals is given in my 'Causes and *if p, even if x, still q*', forthcoming in *Philosophy* 1982, but the point in question does not depend on it.

To press on, it seems clear that it does follow from the fact that E_1 caused E_2 in virtue of a property f of E_1 and property g of E_2 that E_1 and E_2 are in lawlike connection partly or wholly in virtue of properties f and g. If the ground for saying that two events are in some lawlike connection is that they are cause and effect, and it is the case that all of their properties save some residue are irrelevant to their being cause and effect, then they are in the given lawlike connection solely in virtue of that residue of properties. It can be granted not merely that not every true singular statement of causality entails that the events are in lawlike connection under the same descriptions, but also that none does. 'Something weighing two pounds being put on the scale caused the pointer to move to the two-pound mark' does not entail that the events are in lawlike connection under the same descriptions. However, it *does* follow from any statement that the event of the pears' being put on the scale caused the pointer to move to the two-pound mark, and the statement that it did so in virtue of only certain properties, that the events were in lawlike connection by way of those properties. We can call this the Principle of the Nomological Character of Causally-Relevant Properties. It is consistent with and indeed required by any tolerable account of causation and is integral to any account which takes causal relations precisely specified to *be* a species of lawlike relations.[10]

If a mental event causes a physical event, what is the causally-relevant property, or what are the causally-relevant properties, of the mental event? In the case of a physical event which is an action, the mental event for Davidson is, very roughly, a belief and an attitude. More generally, mental events are characterized in terms of what Brentano called intentionality, despite certain problems. Any mental event, however, is identical with a physical event. To speak of a mental event is to speak of an event which also has physical properties. To repeat, then, what is causally-relevant with respect to the mental event?

As noted at the beginning, Davidson remarks that causality is a relation between individual events no matter how described. He goes on to remark that his first claim, the principle of causal interaction between the mental and physical, 'deals with events in extension and is therefore blind to the mental-physical dichotomy'.[11] The first remark, I take it, expresses the truth that we can speak of causal relations between two events however we describe them. In fact we can do better, which is to say we can specify causal relations as holding between the relevant properties of the events. Is the second remark to be taken as denying the proposition that the mental event is a cause in virtue of certain of its properties? If taken in that way, it is surely false. It is not in general mistaken to distinguish causally-relevant properties, and there is no reason to think that mental events are any exception.

---

[10] For an account of the latter kind see my 'Causes and *if p, even if x, still q*', op. cit.

Davidson's account of an action as being caused by a reason, roughly a belief and an attitude, suggests that he takes the mental events in question to be causal as mental. Elsewhere he accepts what can be called the conviction of the efficacy of the mental, 'the efficacy of thought and purpose in the material world'.[12] Again, we have it that Anomalous Monism is not to be confused with 'nothing-but' materialism: 'Conceiving the *Art of the Fugue* was nothing but a complex neural event.'[13] One possible answer to our question, then, is that it is as a mental event that a mental event causes a physical event. It is not a mental event as physical that does the work. Such denials of 'epiphenomenalism' are of course common.

If it is a mental event as mental that causes a physical event, we have an unhappy upshot as soon as we add the truism that a typical physical event said to be caused by a mental event is an effect as a physical event. If we accept the first two of the claims which issue in Anomalous Monism, along with the idea that the mental as mental causes the physical, and the Principle of the Nomological Character of Causally-Relevant Properties, we have the denial of the third claim, that there are no psychophysical lawlike connections. Hence we have a denial of Anomalous Monism. If, on the other hand, we wish to retain the third claim, and accept the idea and the principle just mentioned, we must give up the first claim as we are now understanding it, that there is causal interaction between the mental as mental and the physical.

However, there is also the other possible answer to the question of what is causally-relevant with respect to a mental event. It is the answer that the mental event as physical causes the action. To give this answer is of course to cast a new light on the first claim, that the mental interacts causally with the physical. It becomes the claim that the mental as physical interacts causally with the purely physical. What is important, however, is that the resulting picture seems not to account for a conviction that lies behind acceptance of his first claim when it is naturally understood, as the claim that the mental as mental causes the physical. This is the conviction of the efficacy of the mental, already mentioned. It is the conviction that an event as mental is an ineliminable part of any full explanation of an action. It is the very root of the common denial of various epiphenomenalist doctrines.

---

[11] 'Mental Events', op. cit., p. 215.
[12] 'Mental Events', op. cit., pp. 224–5.
[13] Op. cit., p. 214.

Can this picture of the mind somehow be made tolerable? It cannot be done by a means already noticed, one which depends on the idea that an event would not be the cause it is if it were not a mental as well as a physical event. That will not make the mental character of the event causally relevant with respect to the action, and hence safeguard the conviction of mental efficacy. Nor, evidently, can we gain the end by way of the simple fact that an event can be said to be a cause even when it is picked out by way of a description of its causally irrelevant properties.

The doctrine of supervenience,[14] so far unmentioned, will come to mind. The picture we then get is this: it is a mental event as physical which causes an action; lawlike connection holds between the mental event as physical and the action, but not between the mental event as mental and that same event as physical; however, since the event as mental supervenes on the event as physical, the event as mental is efficacious with respect to the action. The final claim turns on what supervenience comes to, and what it comes to cannot be lawlike connection between the mental and the physical, and is indeed to be understood as no more than the holding of certain universal material conditionals.[15] That is to say that the connection between the mental and the physical is accidental. There is no nomic necessity about the event as physical being the mental event it is. Here, it seems, we do not get the efficacy of the mental.

I have not looked into the question of the truth of the first and third claims out of which Anomalous Monism arises.[16] All I have tried to establish is that the three claims, together with the fact of causally-relevant properties, the principle about their nomological character, and the conviction of the efficacy of the mental, are bad news for Anomalous Monism.

University College London © TED HONDERICH 1982

---

[14] 'Mental Events', op. cit., p. 253. See also 'The Material Mind', op. cit., p. 253.

[15] Clarified by Davidson in discussion.

[16] For arguments against the third claim in particular, see my 'Psychophysical Lawlike Connections and Their Problem' and 'Nomological Dualism: Reply to Four Critics', *Inquiry* 1981, No. 3, No. 4.

INQUIRY

Editor: Alastair Hannay

Vol. 24 (1981), No. 2

On the Philosophical Foundations of Free Logic	Karel Lambert
Dispositions, Explanation, and Behavior	Laird Addis
Must an Opponent of Animal Rights also be an Opponent of Human Rights?	Evelyn B. Pluhar
General Desires as Grounds for the Wrongness of Killing	Brian Smart
Laudan's Pragmatic Alternative to Positivist and Historicist Theories of Science (Larry Laudan: *Progress and Its Problems)*	Gerald Doppelt

Vol. 24 (1981), No. 3

Psychophysical Lawlike Connections and Their Problem	Ted Honderich
Psychophysical Relations	Edgar Wilson
Honderich, Davidson, and the Question of Mental Holism	T. L. S. Sprigge
The Efficacy of Consciousness	J. L. Mackie
On the Relation between Occurrents and Contentful Mental States	Stephen P. Stich
Philosophy without Epistemology (Richard Rorty: *Philosophy and the Mirror of Nature)*	Leon Pompa
Dummett's Prolegomena to any Future Metaphysics (Michael Dummett: *Truth and Other Enigmas)*	J. E. Llewelyn

Subscription price:
Institutions: N.kr 140,– $28.00 – **Individuals N.kr 84,– $17.00.**

Published quarterly by Universitetsforlaget, P.O. Box 2959 Tøyen, Oslo 6, Norway – P.O. Box 258, Irvington-on-Hudson, New York 10533, USA.

MIND

A QUARTERLY REVIEW OF PHILOSOPHY
edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

Vol XCI No 361 January 1982

Articles
- **Events and Objects in Space and Time:** P. M. S. HACKER
- **The 'Scandal' of Cartesian Interactionism:** R. C. RICHARDSON
- **Epistemic Indolence:** R. FOLEY and R. FUMERTON
- **Aristotelian Practical Reason:** M. T. THORNTON
- **Do Zygotes Become People?:** W. R. CARTER

Discussions
- **Existence, Reference and Definite Singular Terms:** L. CHIPMAN
- **Wittgenstein's Debt to Mach's Popular Scientific Lectures:** H. VISSER
- **Intention and Practical Reasoning:** A. B. MILLER
- **A Cognitive Cul-de-Sac:** F. I. DRETSKE
- **Performing an Interpretation; A Reply:** R. A. SHARPE
- **Believing and Affirming:** R. AUDI

Book Reviews

Books Received

Orders to:
Journals Department,
Basil Blackwell Publisher Ltd.,
108 Cowley Road,
Oxford OX4 1JF, England

Annual subscription:
Institutions: £10.00 or $22.50 (Canada $26.00)
Individuals: £8.00 or $18.00 (Canada $21.00)
Single issues: £3.00 or $6.75 (Canada $7.75)

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

PUBLISHER. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. US Mailing Agent, Expediters of the Printed Word Ltd, 527 Madison Avenue, Suite 1217, New York, NY 10022, USA. Second Class postage paid at New York.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £11.00 (inland), £13.50 (overseas), $33.00 (USA), $39.50 (Canada); for individuals £7.00 (inland), £8.60 (overseas), $21.00 (USA), $25.00 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year – in January, March, June and October. Orders should be sent to the publisher.

CONTRIBUTIONS. **Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.**

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham, West Sussex.
Printed and bound in Great Britain by Billing & Sons Limited and Kemp Hall Bindery,
Guildford, London, Oxford, Worcester.

ol. 42 No. 2 (New Series No. 193) March 1982

ANALYSIS

Edited by
Christopher Kirwan

CONTENTS

Average utilitarianisms	T. M. Hurka
Is virtue possible?	Michael Slote
'Five o'clock on the sun': a reply to J. L. Mackie	Jeffrey King
Five o'clock here	Lyle E. Angene
Analysis and the identity of indiscernibles	William Godwin
Rigidity and identity across possible worlds	M. J. More
Intentionality and intuition: a reply to Davies	E. J. Lowe
My body and I—a reply to Fahrnkopf	Michael Welbourne
On the transitivity of epistemic preferability	P. J. R. Millican
Conscious belief	N. M. L. Nathan
The dispositional/categorical distinction	Elizabeth W. Prior
Counting corners correctly	D. H. Mellor
Vanishing powers: a reply to Miller and Wilson	Hillel Steiner
Rowe, self-existence, and the cosmological argument	B. F. Keating
Omnipotence and sin	W. R. Carter
Causal deviancy and multiple intentions	James A. Montmarquet
Imperceptible parts	A. J. Gottlieb

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003-2638

MIND

A QUARTERLY REVIEW
OF PHILOSOPHY
edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

Vol XCI No 362 April 1982

Articles

- **Intentional States of Mechanical Devices:** B. ENÇ
- **On the Need for a Meaning-Theory in a Theory of Meaning:** B. TAYLOR
- **Necessarily Terminating Causal Series and the Contingency Argument:** B. MILLER
- **Dummett's Justification of Deduction:** S. HAACK
- **A Causal Theory of Enjoyment:** W.A. DAVIS

Discussions

- **Ryan on Coercion:** P. WILSON
- **Wittgenstein on Mathematical Proof:** A. AMBROSE
- **Reply to O'Neill on Singular Causal Statements:** G. VISION
- **Frey on Why Animals Cannot Have Simple Desires:** T. REGAN
- **Carney's Proposal:** A.R. MALACHOWSKI
- **Prisoner's Dilemma as an Insoluble Problem:** H. STEINER

Book Reviews

Books Received

Orders to:
Journals Department,
Basil Blackwell Publisher Ltd.,
108 Cowley Road,
Oxford OX4 1JF, England

Annual subscription:
Institutions: £10.00 or $22.50 (Canada $26.00)
Individuals: £8.00 or $18.00 (Canada $21.00)
Single issues: £3.00 or $6.75 (Canada $7.75)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

The Philosophical Quarterly

Special Numbers: The Editors intend to devote an issue in **1982** to articles on **Scientific Realism** and one in **1983** to articles on **Medical Ethics.** Contributions for the latter should be submitted, **by 31 December 1982,** to the Executive Editor, The Philosophical Quarterly, University of St Andrews, Scotland.

Vol. 32 *No.* 126 — *April* 1982

Rigid Designation and Semantic Value	*Colin McGinn*
The Logical Syntax of Deontic Operators	*Alexander Broadie*
Epiphenomenal Qualia	*Frank Jackson*
Law, Morality and Excuses	*J. P. W. Cartwright*
Formal Relations	*Peter Long*

CRITICAL STUDY

Wittgenstein and Experience (L. Wittgenstein, *Remarks on the Philosophy of Psychology*) — *Christopher Peacocke*

BOOK REVIEWS — BOOKS RECEIVED

Annual Subscriptions:
Individuals £8.00 ($20.00) Institutions £16.00 ($40.00)
for orders placed directly with the Journals Manager, Scottish Academic Press Ltd., 33 Montgomery Street, EDINBURGH EH7 5JX, Scotland.

ANALYSIS 42:2 MARCH 1982

AVERAGE UTILITARIANISMS

By T. M. HURKA

AN important class of utilitarian moral theories calculate value, which is what they tell us to maximize, by calculating happiness both across all the persons in human history, and across all the times in human history.[1] Classical utilitarianism is one of the theories in this class, and it uses summative principles to make both these calculations. It determines the happiness felt by a group of *n* persons by summing the happiness felt by all of its members; and it determines the happiness felt in a period of *n* years by summing the happiness felt in all of those years. Nor does it matter what order classical utilitarianism applies these two summative principles in. If it wants, it can begin by summing the happiness felt by a single person at all the times in his life; this will give it a measure of the total happiness in his life, and it can then sum these measures across lives. Or it can begin by summing the happiness felt by all the persons in the world at a single time; this will give it a measure of the total happiness in the world at that time, and it can then sum these measures across times. Whatever order classical utilitarianism applies its two principles in the result will be exactly the same; and as a result nothing turns on the order it actually adopts.

Average utilitarian theories use averaging principles at least somewhere in their calculations of happiness. But there are several such theories in the class of those which calculate across all the persons and times in human history, and in all of them a great deal turns on the order in which the principles in them are applied. Let us begin by considering the average utilitarian theories in this class which use an averaging principle to calculate happiness across persons, and a summative principle to calculate happiness across times (they are probably closest to what philosophers have had in mind when they have spoken about "average utilitarianism" in the singular). These theories can begin by summing the happiness felt by a single person at all the times in his life; this will give them a measure of the total happiness in his life, and they can average these measures across lives. Or they can begin by calculating the average amount of happiness felt by each person alive at a single moment; this will give them a measure of the average happiness per person at that moment, and they can then sum these measures across moments. But these two procedures are not guaranteed to yield the same result. There are circumstances in which they yield different results, and we must therefore regard the theories which adopt them as different average utilitarian theories. Let us call the theory which adopts the first procedure—the one which first sums across times and then averages across persons—A1, and the theory which first averages across persons and then sums across times A2.

---

[1] Utilitarians often include animal happiness in their calculations of value. I have only excluded it because this simplifies some of my arguments considerably.

Both A1 and A2 have unattractive consequences, but these unattractive consequences are different. A2 requires us to kill people who will feel less than the average amount of happiness from now on into the future. Let us imagine that the world is an extremely happy place, in which almost everyone is enjoying 1000 units of happiness every year (this on a scale where the best we do today is 10 units of happiness a year). But one person is feeling only 900 units of happiness every year. Then even though this person's life is a very happy one A2 requires us to kill him. His existence at any future moment would lower the average happiness per person at that moment, and we must therefore prevent it. The fact that A2 requires us to kill below-average people in this way surely constitutes a powerful objection against it; but it is important to realize that the same objection cannot be raised against A1. A1 tells us to maximize the average happiness per human life lived; and it calculates happiness within lives using a summative principle. To end the life of a man who is enjoying 900 units of happiness every year is to decrease the amount of happiness his life will contain as a whole, and it is therefore also to decrease the contribution his life will make to the general average. Killing him will not remove his life from the list of those lived in human history, and A1 therefore has no reason to recommend it.[2]

Although A1 does not have unattractive consequences about killing it does have unattractive consequences in other areas. Imagine that human lives up to this point in history have been extremely happy, containing an average of 100,000 units of happiness per life lived. But from now on human lives will be slightly less happy. As a result of unfortunate climatic changes (winter has disappeared and people have to live all year long without ice and snow) human lives will contain only 90,000 units of happiness. Then A1 requires us all to stop having children, and indeed to let the human race die out. Even though the children we produced would lead very happy lies (90,000 units of happiness where the best we do today is 1000), these lives would lower the average happiness per life in human history, and they must therefore not be lived. The fact that A1 requires us to let the human race die out in this situation surely constitutes a powerful objection against it; but it is important to realize that the same objection cannot be raised against A2. A2 calculates the average happiness per person at a single moment, but it then goes on to sum these averages across moments. To prevent the existence of people who would feel 90,000 units of happiness per person each year is to decrease the average happiness at every moment in the future from a large positive number to 0; and it is therefore also to decrease the sum of those averages across all the moments in human history.

---

[2] The claim that "average utilitarianism" (in the singular) requires the killing of below-average people has been made by a number of writers, among them James E. Meade, Richard Henson, Robert Nozick, Philip E. Devine, Peter Singer, Robert Scott, Jr., and Wayne Sumner. It is not clear, however, that these writers had A2 in mind when they made this claim, for some average utilitarian theories which calculate happiness across only *future* persons and times also require the killing of below-average people, though who counts as below-average in these theories is slightly different. Claims about the consequences of "average utilitarianism" in the singular have not always been made, it seems to me, with a full awareness of the large number of different average utilitarian theories there are.

A1 and A2 also have unattractive consequences associated with their use of summative principles across times (some of these were pointed out to me by Derek Parfit). A2, for instance, entails something like the "repugnant conclusion" associated with classical utilitarianism. Let us try to imagine an ideal (finite) course of human history, one in which the average happiness per person is very high for a very long (finite) period of time. Then if A2 is correct there is another vastly longer course of human history which is better than this one, even though the average happiness per person in this second course is never more than barely positive, and the experiences of the people in it never more than barely worth having. The second course is better because, if it is long enough, the sum of the average happiness per person at each time is greater than it is in the first course. A1 sometimes shares this unattractive consequence with A2 and sometimes does not. If the lives lived in the second course of history are also vastly longer than those lived in the first course, and therefore contain more total happiness than those in the first course, A1 joins A2 in preferring the longer second course. But if the lives in the second course are roughly the same length as those in the first course, A1 prefers the shorter first course. A1 also entails something like the repugnant conclusion. Let us try to imagine an ideal human life (of finite duration), one in which the average happiness per moment is very high for a very long (finite) period of time. Then if A1 is correct there is another vastly longer human life which it would be better to have lived than this one, even though the average happiness per moment in this second life is never more than barely positive, and the experiences it contains never more than barely worth having. A2 sometimes shares this unattractive consequence with A1 and sometimes does not. If the other lives lived alongside these two lives have an average happiness per moment which is lower than or equal to that of the second life, or if there are no other lives lived alongside the two lives, A2 joins A1 in preferring the longer second life. But if the other lives lived alongside these two have a higher average happiness per moment than the second, A2 prefers the shorter first life.

A1 and A2, then, are open to a number of objections on account of their consequences for particular moral choices. A more theoretical objection, and one which applies to both theories equally, is suggested by a famous passage in Henry Sidgwick's *The Methods of Ethics* (7th edition, Macmillan 1907, p. 418). In this passage Sidgwick asks how an egoistic hedonist can justify giving equal weight to pleasures felt at different times in his own life and no weight to pleasures felt by different persons, when the problems raised by pleasures felt at different times and pleasures felt by different persons are from a formal point of view exactly analogous. The question Sidgwick asks about weightings can also be asked about calculations of happiness. If the problems raised by calculating happiness across persons and times are also formally analogous, as they clearly are, how can A1 and A2 justify using averaging principles to calculate happiness across persons and summative principles to calculate happiness across times? If A1 and A2 could point to some relevant difference between persons and times then of course their use of these different principles would be beyond reproach; but it is difficult to see what this relevant difference could be, and difficult to see how their use of different principles could ever be justified.

Reflection on this "symmetry objection" leads us to consider the average utilitarian theories in our class which are not open to it, because they use averaging principles both across persons and across times. There are two such theories, differing from each other in the order in which their two averaging principles are applied. The first of these theories (call it A3) begins as A1 did by calculating the happiness within a single human life, but instead of calculating the total amount of happiness in a life A3 calculates the average happiness per moment in it. Having calculated this average, however, it then follows A1 in averaging across lives, and tells us to maximize the average of the average happiness per moment in all the lives in human history. The second theory (call it A4) begins as A2 did by calculating the average happiness per person at a single moment; but instead of telling us to maximize the sum of these averages across moments it tells us to maximize their average.

Although A3 and A4 are not open to the symmetry objection, and do not entail anything like the repugnant conclusion, their consequences for other particular moral choices are even more unattractive than those of A1 and A2. A3 has the same unattractive consequences about letting the human race die out that A1 has, but it also has some unattractive consequences about killing. Imagine that a person has had an extremely happy life up to the present, enjoying 1000 units of happiness in every year that he has lived. But in the future he will enjoy only 900 units of happiness every year. Then even though these extra years would be very happy ones A3 says they must not be lived. They would lower the average happiness per moment in this person's life, and we must therefore kill him. (To see that A4 is distinct from A3 imagine that this person, who has been happier than everyone else in the past, will still be happier than everyone else in the future, even though everyone else will be happier in the future than they were in the past. In this situation A4 says he must be kept alive.) A4 has the same unattractive consequences about killing that A2 has, but it also has some unattractive consequences about letting the human race die out. Imagine that the average happiness per person has been very high throughout human history, with every person enjoying at least 1000 units of happiness in every year of his life. But recent climatic changes mean that in the future everyone will enjoy only 900 units of happiness every year. Then even though people would be very happy in the future A4 requires us to let the human race die out. The average average happiness per person would be lowered if there were any other moments at which humans were alive, and procreation must therefore come to an end.

The last average utilitarian theories in our class for us to consider (and there are once again two of them, which we can call A5 and A6) use summative principles across persons and averaging principles across times. A5 and A6 are probably the least attractive average utilitarian theories of all in our class. They are open to the symmetry objection, and combine many of the most unattractive consequences of classical utilitarianism and A1 and A2. The identification of these consequences I will leave to the reader.

"Average utilitarianism" is usually discussed and criticized as if it were a single theory. This is a mistake. There are six different theories which can be called average utilitarian in the class of those which calculate happiness across all the persons and times in human history alone, and although these theories are all open to powerful objections the objections are in many cases importantly different.

University of Calgary

© T. M. HURKA 1982

IS VIRTUE POSSIBLE?

By MICHAEL SLOTE

MUCH attention has recently been paid to the nature and interrelations of the moral virtues and to the psychology of their embodiment in particular individuals. Although it has typically been assumed that a virtuous individual knows what is right in different situations and acts accordingly, both Iris Murdoch in *The Sovereignty of Good* (London: Routledge and Kegan Paul, 1970) and John McDowell in 'Virtue and Reason' (*The Monist* 62, 1979, pp. 331–50) have laid particular stress on the (Aristotelian) idea of perception in their accounts of the epistemology of virtue. Murdoch and McDowell offer a marvellously subtle and complex treatment of virtue, and their ideas have rich reverberations for our understanding of Plato and Aristotle's ethical thought. But their account of moral virtue is also importantly incomplete. And in what follows I shall argue, in particular, that their exclusive focus on internal defects that prevent people from seeing and doing what is right makes them overestimate the general accessibility of moral virtue and ignore the historical dimension of our attempts to understand "moral reality."[1]

I

The virtuous individual is often said to be capable of knowing right from wrong in the wide variety of circumstances that might arise in his life. This knowledge may not be embodied in any (complex) ethical precept completely formulable by him or even in principle; indeed that is part of the reason for saying, with Aristotle, that overall virtue—or any particular virtue—involves (or is) a quasi-perceptual capacity for recognizing what, in the various circumstances in which one may be called upon to act, is the right thing to do. But knowledge is nonetheless required if one is to be a virtuous person and not just someone who by benign accident or at the behest of others does what a virtuous individual is supposed to do from knowledge.

Most accounts of moral virtue treat it as a disposition to be acquired only through a proper upbringing and education. But McDowell and Murdoch place a special emphasis on another precondition of virtue. What prevents some people from being virtuous at a given time can be something about them at that time which, though it may to some extent be due to faulty training, has independent explanatory force in understanding their lack of virtue. A person can be too self-centred, too emotional, too proud to get things morally right, and because such individual defects are widespread and difficult to overcome, Murdoch and McDowell both speak of (overall) virtue as very difficult to attain.

---

[1] I shall be responding to a composite (and partial) picture of what Murdoch and McDowell say, but shall note differences of emphasis where relevant. McDowell acknowledges a debt to Murdoch and at no point indicates any disagreement with her work. Note further that my criticism will not hinge on anything either of them says, or fails to say, about *akrasia*.

Nonetheless, I am more struck by the possibilities of virtue McDowell and Murdoch do not rule out than by those they do. For by talking about what virtuous individuals do and are like, they both imply that there *have been* virtuous individuals, and Murdoch makes an explicit commitment to that assumption by claiming that moral virtue can be found not only in the likes of Socrates, but among peasants as well. Furthermore, Aristotle (see, e.g., *Nicomachean Ethics* 1143a 25–35; 1144b 30–1145a 6) appears to have believed in the existence of virtuous individuals in his own day, and McDowell not only says he wishes to defend an Aristotelian conception of the inculcation and practice of virtue, but mentions no factor that might have made it more difficult to acquire virtue in Aristotle's time than in our own.

Indeed, if the sorts of factors they mention were the sole impediments to the development and exercise of virtue, then the attainment of virtue, as traditionally conceived, might be no more problematic than McDowell and Murdoch suggest. But it seems to me that there are other, less individualistic factors that must be brought into an account of virtue, and in the light of these I think it doubtful that virtue existed in the ancient world or can ever be found among the uneducated. It even becomes questionable whether virtue is attainable today, but these difficulties are perhaps best formulated in terms of examples.

II

We now regard slavery as an unjust institution that offends the basic rights of the enslaved. But in the ancient world—in Greece, in Rome, among the Church Fathers—no one ever wrote, or is known to have spoken, against slavery as such.[2] There were occasional discussions of the abuses and justifications of slavery, but it was never held that owning slaves was simply wrong. Murdoch and McDowell say that the virtuous individual takes account of the rights of others and of what is just and fair, and they use such moral notions *tout court*, with all the objectivity that that implies. But if we agree that slavery offends the rights of slaves, can we plausibly avoid saying that the keeping, buying, and selling of slaves not only is but was wrong; and will not that in turn commit us to saying that those engaged in such practices were not morally virtuous individuals? But such criticism will then also naturally extend beyond those who actually possessed slaves to people too poor to own slaves, perhaps even to slaves themselves. Even those who could not afford slaves would presumably have been *disposed* to own slaves if good fortune had made it possible. For slavery was simply accepted, as a natural fact, in the ancient world. No other system was known or, perhaps, even thought possible. So with benefit of hindsight, the assumption that slavery is wrong can lead one to question Aristotle's assumption that there actually were virtuous individuals to be found in his society and McDowell's seemingly unqualified endorsement of Aristotle's views about the nature and realization of virtue.

---

[2] See M. I. Finley, *Ancient Slavery and Modern Ideology*, London: Chatto and Windus, 1980, pp. 120f. Finley mentions only Euripides as a possible exception to the above generalisation.

McDowell and Murdoch say that egotism, wilfulness, laziness, and bad temper can render someone incapable of seeing all that the virtuous person must see. But such variable factors of individual psychology—and various deficiencies of training—are the only barriers to virtue they mention, and if nothing else could stand in the way of virtue, then there would perhaps be no reason why virtue should not have been present in the sorts of individuals Aristotle regarded as virtuous. When, however, we consider the example of slavery—and other examples to be mentioned shortly—it can begin to seem questionable whether the above factors represent the only bars to moral virtue. If virtue requires the disposition to act justly, do the right thing, with regard to important issues, can we sensibly suppose that the failure of the ancient world to attain and act upon a correct moral view of slavery merely reflects personal weaknesses or deficient methods of moral training that, unluckily, happened to be fairly universal in those days? Presumably not.

I have never heard anyone suggest that the Greeks, ancient Hebrews, or early Christians were less able to instil moral values into their children than we are today—there is reason, in fact, to believe the contrary. And to say that the ancients were blinded to the injustice of slavery (say) by their inordinate greed would not account for the fact that those too poor to own slaves (even, I suppose, the large majority of slaves themselves) never thought of questioning slavery. Nor was ancient belief in the naturalness and inevitability of slavery (and in the existence of "natural slaves") *merely* a widespread rationalization of individual avarice. Rather, that belief is at least partly to be explained by the fact that slavery was a universal phenomenon. Just as ignorance of the alternative terms used by other languages can make matters of linguistic convention seem to be inevitable facts of nature, so too can ignorance of alternatives to a given social arrangement instil the belief that that arrangement is natural and inevitable and thus beyond the possibility of radical moral criticism. So if the ancients were unable to see what virtue required in regard to slavery, that was not due to personal limitations (alone) but requires some explanation by social and historical forces, by cultural limitations, if you will.[3] And if we today can see the wrongness of slavery, that is in part because we have the benefit of knowledge that makes slavery seem less natural and inevitable. For unlike the ancients we know of "experiments" in living without slavery; we possess an historical record of societies, among them our own, where slavery has been absent and people have survived and flourished, nonetheless.

Similar points arise in connection with issues other than slavery. Murdoch insists (p. 1f.) that an "unexamined life" can be virtuous and that there are and can be virtuous peasants. And I believe she makes such claims because she regards personal defects as the only barriers to correct moral vision, to virtue. But there is reason to wonder whether such factors are the only ones that affect whether a peasant does the right thing. Consider, for example, a peasant father or mother who refuses to help his/her daughter leave a husband who is clearly unloving and abusive. We believe, do we not, that a woman has the right to leave such a husband (though there may be wrong *ways* of doing so), but what if the woman's own parents refuse to help her in this, telling her to stay with her husband and even rendering it impossible for her to leave him (she needs a small amount of money from them, say, in order to get away)? Such actions presumably express a failure to acknowledge the rights of the unhappy wife, and most of us would say that they are a wrong way to act towards a daughter in distress and in need. But what prevents her parents from doing right by her? Need it be envy of the new life she hopes to lead and they never had or (in the case of the father) sexist antagonism towards women generally? Need it be self-righteous anger reflecting frustration at having chosen wrongly for her (we are talking about a peasant society) or selfish discontent at having her once more dependent upon them? Can it not simply be that, despite all their love and desire to help her, they believe in *a holy book* that commands all women to stay with their husbands through thick and thin? Murdoch herself (pp. 74, 101) says that traditional religion is no barrier—is even, indeed, a positive aid—to virtue and right action. But then what can be said about the case at hand? Unless we hold, implausibly and against Murdoch's own tendencies, that only a blind or selfish person would accept the holiness of the book in question, we shall have to grant that the peasant parents may do the wrong thing for reasons having to do with social and cultural influences (as reflected of course in them) rather than through the sorts of personal, individually variable, vision-preventing defects she so exclusively focuses upon. And there are many, many other such situations where tradition and milieu would put "enlightened" views of right and wrong beyond the reach of (devout) peasants. So if Murdoch is right to insist that right action is a criterion of virtue and if modern views of the rights of women are correct, then the existence of virtue in unexamined lives seem very dubious.[4]

---

[3] On the distinction between individual-psychological and social/historical factors, cf. C. Ryan, "Beyond Beliefs," *American Philosophical Quarterly* 18, 1981, pp. 33–41 and references therein. On the unavailability to the Greeks of "enlightened" moral views about slavery, see T. H. Green, *Prolegomena to Ethics*, Oxford: 1883, ch. 5.

Nor would it be very easy for Murdoch and McDowell to evade the above conclusions by claiming that slavery was *not* wrong in the special circumstances of the ancient world or that "women's rights" *do not exist* in peasant societies. For they are self-avowed moral realists who hold that moral beings are seeking to understand an independent moral reality and, where appropriate (e.g., Murdoch, pp. 55, 59, 69; McDowell, pp. 346ff.), they have sought to distinguish their views from any form of moral relativism. Since, in addition, Murdoch and McDowell both speak freely, and in the most objectivistic terms, of justice, fairness, and rights, I find it impossible to imagine either of them being happy with the view that slavery was fair in ancient Greece or with the view that (certain) peasant women have no right to leave their husbands.

III

In that case a very real question must arise about whether overall virtue is possible even for us. We have been speaking as if the wide-ranging capacity for right action that seems problematic in relation to the Greeks or peasant societies presented no particular problem for such modern individuals as can perceive the rights of slaves and women and with proper education overcome the individual-psychological barriers to virtue that McDowell and Murdoch describe. But in fact I have no intention of being thus smug about current values and conceptions of morality. I have, for example, harped on the wrongness of slavery and the unavailability of that moral fact, and others, to the ancients, but I too admire Greek moral philosophy, and that very admiration causes me to wonder whether virtue is any more accessible nowadays than it was in classical antiquity. I assume that we now see aspects of right and wrong that the ancient world missed; but may we not also fail to see things the Greeks were able to see, and, more significantly, may there not be other moral issues on which both the ancients and we ourselves are blind? Someday, our descendants may regard twentieth century morality and moral philosophy as having been as hopelessly wrongheaded on certain issues (vegetarianism? children's rights? patients' rights?) as were ancient views about slavery. And they may have as much reason to do so as we currently have to be critical of ancient attitudes to slavery.

A natural enough induction from the history of previous *scientific* theories tells us that today's theories are very likely to turn out to be

[4] Murdoch says that no one is morally perfect. But my assumption that a virtuous individual should be capable of right action with regard to such important issues as slavery is hardly an insistence on moral *perfection*.

mistaken in important ways. Is there any less reason to believe the same about current-day moral thinking? If not, then there is reason to believe that moral virtue as traditionally understood is not accessible even today, but rather, like various kinds of scientific truth, is to be attained, if at all, only at the historical limit of human cultural endeavour, in a long run that no individual may ever encompass. McDowell and Murdoch have underscored the individually understandable obstacles to virtue, but because moral virtue is subject to social/historical influences and limitations that they never mention, the very possibility of virtue comes into question.

I believe, however, that I can see one reason why Murdoch and McDowell limit themselves to individualistic factors in talking about virtue and neglect the larger cultural factors I have been emphasizing. Both are wary of the scientism that holds that science can tell us everything about reality and that what it doesn't speak of doesn't exist. And they deliberately stay clear of those analogies between moral philosophy and science that moral philosophers are often so eager to develop. One such attempted analogy can be found in the idea of moral philosophy as a progressive discipline that (occasionally) arrives at new and better moral knowledge (theories) through the historical accumulation of greater non-moral knowledge—and by building on the ideas of previous (moral) philosophers. In rejecting such analogies, Murdoch and McDowell limit themselves to focusing on ontogenetic impediments to moral knowledge (vision) and right action: such limits and impediments as are due to be overcome by means of the phylogenesis of moral knowledge could be acknowledged or featured only by admitting the sort of analogy between progressing science and moral philosophy that seems repugnant to Murdoch and McDowell. [Thus when Murdoch speaks (p. 38) of the "historically conditioned" aspects of virtue, she is in fact speaking only of the history of a particular individual who is to become virtuous by overcoming his own passions and foibles. History in the usual sense remains out of the picture.]

It is clear that I am not altogether averse to the above analogy between moral philosophy (and moral ideas generally) and science. It is dangerous to speak of moral progress, but I have risked doing so with regard to certain specific issues like slavery and the right (in principle) of a woman to leave her husband. Nonetheless, the main focus of my acceptance of the analogy is not the idea of similar progress, but rather of similar historical character. There seems to be something historically *developmental* about moral philosophy, even if this does not amount to progress as science understands (and achieves) it. Such an analogy with science does not force us to the sort of scientism McDowell and Murdoch decry. For one could hold that there was (phylogenetic) historical progress in moral theory without assuming that that occurred through some sort of convergence with scientific inquiry; the progress, if it occurs, can be directed, if you like, to a moral reality of the sort Murdoch and McDowell envisage.

In rejecting the analogy with historically developing science, Murdoch (pp. 27, 76) and McDowell (pp. 346ff.) also imply that moral philosophers have no contribution to make to our understanding (knowledge) of right and wrong—they can clarify concepts and indicate the structures and processes of the moral life, but that is another matter. This, however, ignores, or perhaps rejects, some important aspects of the history of moral philosophy. When it was first discussed in the eighteenth century, Utilitarianism was an original moral idea. No one had seriously thought that right action consisted in just what Utilitarianism says it does. Similarly, to take a more recent example, Rawls's difference principle (as opposed to aspects of his argument for it) seems to be a totally new substantive moral conception. Now even if neither the principle of utility nor the difference principle is actually true (or thus represents the moral knowledge that the ancients lacked and needed for the attainment of virtue), they are ideas worth taking seriously, real contributions to the attempt to know what is right (think how much has been learned by attempts to *refute* Utilitarianism and Rawls). And if new and promising ideas not about moral concepts but about actual right and wrong can thus emerge through the individual insight or vision of particular moral philosophers (or others), does this not locate another reason why virtue may not be attainable in those too-often-neglected factors of historical development that make some sort of analogy between science and moral thinking entirely plausible?

By focusing on characteristics of virtuous individuals, McDowell seeks to give an account of morality, as he says, from the inside out (p. 331). Moral principles become secondary or drop out altogether, because the tendencies to right action that virtue requires cannot, he thinks, be grasped or formulated in any statable general moral principle(s). But if, as I have suggested, the development of moral thought and the realization of virtue wait upon the thinking up and theoretical sifting of such moral principles as the Utilitarians and Rawls introduced, then there is a place for an independent "principles" approach to moral philosophy alongside efforts to conceive morality from the standpoint of the virtuous individual.[5]

Trinity College, Dublin © MICHAEL SLOTE 1982

---

[5] I am indebted to John McDowell, Myles Burnyeat, John Dillon, and Edward Erwin for helpful suggestions.

'FIVE O'CLOCK ON THE SUN': A REPLY TO J. L. MACKIE

By JEFFREY KING

IN his recent paper 'Five O'clock On The Sun' (ANALYSIS 41.3, June 1981) J. L. Mackie makes clear with admirable precision what he thinks must be added to the remark 'It's 5 o'clock on the sun' to, as he puts it, 'give it a clear meaning and a determinate truth value'. While I do have certain misgivings regarding Mr. Mackie's explanation of what might be meant by 'It's 5 o'clock on the sun', I shall not pursue these here. For it seems to me that whether Mackie's explanation of 'It's 5 o'clock on the sun' be cogent or not, he has missed the point of Wittgenstein's remark in section 350 of *Philosophical Investigations.*

As is clear from the passage itself, what Wittgenstein is objecting to is a particular explanation of 'It's 5 o'clock on the sun'. Specifically, Wittgenstein objects to explaining 'It's 5 o'clock on the sun' by saying that it means that it is the same time on the sun as it is here when it is five o'clock here. As Wittgenstein quite rightly points out, this "explanation" does not tell us anything we don't already know. I know that if someone says 'It's 5 o'clock on the sun' part of what is being asserted is that it is the same time there as it is here when it is five o'clock here. What I still don't know, however, is in what cases it is the same time there as it is here when it is five o'clock here.

Suppose someone says to me 'My dog is laughing' and, seeing the bewildered look on my face, explains by saying 'I mean that my dog is doing the same thing that you do when you laugh'. Clearly, such an "explanation" doesn't tell me anything I don't already know, and thus is of no use. I already know that part of what was being asserted was that this person's dog was doing the same thing that I do when I laugh. And it is for precisely this reason that this "explanation" is useless here.

Now it is this type of empty explanation that Wittgenstein is objecting to in section 350. Wittgenstein's point is *not* that an explanation of 'It's 5 o'clock on the sun' cannot be given at all (though this too may be true), but that the particular (alleged) explanation 'It means that it is the same time on the sun as it is here when it is 5 o'clock here' explains nothing. Mr. Mackie, however, seems to have misunderstood this point. For Mr. Mackie's explanation of what might be meant by 'It's 5 o'clock on the sun' is a different explanation from the one criticized by Wittgenstein in section 350. That is, Mackie's explanation seems to me to amount to considerably more than 'It means that it is the same time on the sun as it is here when it is 5 o'clock here.'

University of California at San Diego

© JEFFREY KING 1982

FIVE O'CLOCK HERE

By LYLE E. ANGENE

> "But if I suppose that someone has a pain, then I am simply supposing that he has just the same as I have so often had."—That gets us no further. It is as if I were to say: "You surely know what 'It is 5 o'clock here' means; so you also know what 'It's 5 o'clock on the sun' means. It means simply that it is just the same time there as it is here when it is 5 o'clock."—The explanation by means of *identity* does not work here. For I know well enough that one can call 5 o'clock here and 5 o'clock there "the same time", but what I do not know is in what cases one is to speak of its being the same time here and there.
>
> (*Philosophical Investigations* I 350)

AFTER quoting this passage, J. L. Mackie (ANALYSIS 41.3, June 1981) claims: 'Wittgenstein is here appealing to some difficulty in understanding "It's 5 o'clock on the sun" as evidence that one could not derive an understanding of "He has a pain" directly by simple analogy, from a presumed prior understanding of "I have a pain".' Mackie suggests that the remark 'It's 5 o'clock on the sun' is not really that problematic, and he offers three ways of attaching three different senses to it. I suggest that Mackie has looked in the wrong place for the wrong difficulty and hence has failed to understand or undermine the passage he quotes.

The remark with which Wittgenstein takes issue is: 'It means simply that it is just the same time there as it is here when it is 5 o'clock [here][1].' To see what is wrongheaded about this remark, contrast it with the following. 'To say that it is 5 o'clock in London just means that it is the same time of day in London as that time of day which we here in Virginia call 5 o'clock.' Because the notion of a day in London is unproblematic and because I know what time of day is 5 p.m. here (late afternoon, near evening, quitting time, almost supper time), I can easily construe this remark as a claim about what time (hour, part of a day) 5 p.m. in London is.

But now the difficulty in saying of the sun, 'It is the same time there as it is here when it is 5 o'clock here' should be clear. The difficulty has nothing to do with simultaneity, whether distant or otherwise, for simultaneity is neither being asserted or denied. The problem is simply that there is no "time of day" on the sun, for there are no days, nights, afternoons, or evenings there. Nor is it clear how we are to give sense to talk about lunch time, bed time, quitting time, or noon time on the sun. The natural and social rhythms of daily life that permeate our sense of what time it is *here* are not to be found on the sun.

---

[1] Anscombe's translation omits this last 'here', but '*hier*' is present in the German, and, as I shall show, crucial to the sense of the passage.

Wittgenstein is not asserting that the phrase '5 o'clock on the sun' has no sense or that it cannot be given sense but just that it cannot be given sense in a particular way. 'The explanation by means of identity does not work here.' As Mackie shows, one can stipulate what is to count as '5 o'clock on the sun', but this does not affect Wittgenstein's point that this cannot be done by saying that it is the same time *there* as it is *here* when it is 5 o'clock *here*.

Perhaps Wittgenstein's illustration fails to work for some because they overlook the fact that it makes sense to ask what time 5 o'clock here is. Such oversight is perhaps natural since, though most of us know what time 5 o'clock here is, I am unaware of any satisfactory explication of what it is we know and how we come to know it.

As I understand it then, Wittgenstein's illustration is supposed to show that it is unhelpful to announce that an expression used "there" means "there" what it means "here" when used "here" *if* those features which give the expression sense "here" are not present "there". Applied in the case of pain, the point would be that I can't give the expression 'pain in the stove' sense by saying pain in the stove is *in the stove* what pain in me is *in me*. And *if* my account of my pain is given in terms essentially unique to me, then I can't just stipulate that pain in you is *in you* what pain in me is *in me*. There may be difficulties with this view, but Mackie's suggestions about ways of giving sense to talk about time on the sun are irrelevant to the point Wittgenstein was attempting to illustrate and to the appropriateness of the example he chose to illustrate it.

The College of William and Mary

© LYLE E. ANGENE 1982.

ANALYSIS AND THE IDENTITY OF INDISCERNIBLES

By WILLIAM GODWIN

THERE seems to be an incoherence at the heart of the *Tractatus*, arising from certain requirements of its theory of analysis on the one hand, and its rejection of the principle of the identity of indiscernibles on the other. I shall give what seems to me to be the argument for this, and then outline its significance.

I

At several places in the *Tractatus* and *Notebooks* Wittgenstein insists that every ordinary proposition must in principle be analysable into a truth-function of elementary propositions (4.211, 3.251, 3.2–3.201, 2.0201[1]; *Notebooks* (NB), ed. Anscombe, pp. 63 (b), 62). The possibility of propositions or sense required the possibility of elementary propositions, since sense had to be completely determinate (3.14, 3.23; NB pp. 23(e), 61(f)), and elementary propositions simply expressed this determinacy explicitly: only with them did the objects meant by every proposition stand in a one-to-one relation with propositional elements; no other type of proposition consisted solely of terms themselves meaning the extra-linguistic meanings of every proposition (3.2–3.221, 4.21–4.22, 4.23–4.24). In as much as the sense of every proposition was therefore identical with that of some truth-function of elementary propositions (4.51, 4.52), every proposition was in principle analysable into or expressible as such a truth-function (4.221, 3.2–3.201, 3.25, 3.3442; NB p. 60(c)).

Although sense had to be absolutely determinate, ordinary or non-elementary propositions were thus in a sense indeterminate: not being composed of 'names', they did not themselves reveal which particular configurations of objects they meant. This too is identified with their generality and their possibility of expression as generalisations (5.526, 3.24, 2.0201; NB pp. 69, 46, 43; *Notes on Logic* II(L); *Letters to Russell, Keynes and Moore*, ed. v. Wright, pp. 43–4; *Philosophical Grammar*, ed. Rhees, p. 211). Analysis then culminates in an inference from some such general proposition to a particular truth-function of elementary propositions (see above references, but especially 5.526(b)).

II

The possibility of inference of truth-function of elementary propositions from general proposition—required if analysis is even in principle completable—evidently entails a certain proviso. It requires that, for any given state of affairs S, and for any pair of objects *a* and *b* such that *a* is a constituent of S, it must be impossible to replace *a* by *b* in S; that is, *b* cannot combine in just the same way in S as *a* can. For suppose that this were possible; then there would always be more than one truth-function of elementary propositions we could form out of the general proposition: it would never be *impossible* for the general proposition to be true and the truth-function of elementary propositions to be false, since, for any such truth-function we put in place of the general proposition in trying to arrive at a complete analysis, there could always be some other such truth-function, identical perhaps in every way to ours save in containing the name of some object other than one of the objects named in our original elementary proposition.[2] In order to secure the possibility of complete analysis, then, it must be impossible for any two objects to combine in exactly the same way in otherwise identical states of affairs; or, equivalently, for any two names of objects to concatenate in just the same way with other names to form elementary propositions which would otherwise be identical.

---

[1] All decimal-point references are to the *Tractatus*.

This proviso clearly entails the principle of the identity of indiscernibles, for if no two objects can configure in just the same way then no two objects can have all their properties in common. But the principle is explicitly rejected by Wittgenstein at 5.5302. He there insists that 'two objects have all their properties in common' expresses a proposition with sense, and this means, for him, that the issue is an empirical or contingent one, and not a matter of any sort of necessity, since only contingent or empirical propositions have *sense*, or are 'propositions' in the strict sense of the word (2.1–3.1). Still another way of putting it is to say that only what is empirical or contingent can be said, that what is 'logical' or 'formal' is unsayable, the attempt to say it producing nonsense.

III

Indeed it must for Wittgenstein be sayable, or contingent, that 'two objects have all their properties in common' (or, correspondingly, that they do not), for at least the following reason. First, some of an object's properties naturally relate to its actual concatenations with other objects (their 'material properties'—2.0231), and on the picture theory these actual concatenations are precisely the senses of propositions, rendering their 'sayability' or contingency internal to that theory; but, secondly, if it is thus sayable that such-and-such objects are so-and-so configured, if, that is, this is a fact, or state of affairs (*Sachverhalt*—2.0272, 2.034–2.04) then surely it is equally sayable, equally a fact, that such-and-such objects are actually configured in different ways or have different properties from such-and-such other objects (or that they are identically configured). But if this is a fact, if it is sayable, or has sense, then the identity of indiscernibles must be rejected; it can only have an empirical analogue—'no two objects have all their properties in common', where there is no implicit 'can'.[3]

---

[2] It might be objected at this stage that, strictly speaking, Wittgenstein allows the predicate 'true' to apply only to genuine 'pictures of reality'—at 4.06, for example, he explicitly says 'A proposition can be true or false only in virtue of being a picture of reality'—and that generalisations or 'protopictures' are by definition not pictures of reality (3.315). The crux of this issue is whether we are to allow propositions other than elementary ones to enter into inferential relations. If we do, and the requirement of analysis together, for example, with the description of mathematical method (6.2341–6.241), requires that we do, then we must either allow 'true' to apply to other types of significant proposition or try to define the notions of inference and logical equivalence in a way which makes no use whatever of the notion of truth. Wittgenstein himself certainly attempted no such thing, and it is very hard to see how it could be.

So there seems to be an incoherence at the heart of the *Tractatus*, in the 'picture theory' itself. The theory requires the possibility of analytically arriving at (required) elementary propositions which entails a proviso which in turn entails the identity of indiscernibles; but the identity of indiscernibles must be rejected on the theory because the theory requires that 'no two objects have all their properties in common' express a contingent or empirical proposition, and not any kind of necessity.

The importance of the incoherence seems to be twofold. First of all, it constitutes a refutation of perhaps the most rigorous development of a certain pervasive conception of language. Wittgenstein in his later work is well known to have repudiated his *Tractatus* conception of meaning as false or misleading; it is now possible to repudiate it as actually incoherent. But secondly, it ought to prompt us into searching out and revealing those features of the underlying conception which produce the incoherence. The incoherence arises only against the background of a certain conception of philosophical analysis: what features of the 'picture theory' therefore give rise to that conception of analysis? One broad direction in which to look would be at the kind of conception of the relation between meaning and the notion of speakers' understanding which the theory embodies, and which must presumably, in one way or another, inform any theory with a comparable notion of what analysis is and what its aims are. And once we start to focus on such features the incoherence identified above will itself become much less significant.

Oriel College, Oxford © William Godwin 1982

---

[3] We must not suppose that the above point is at all comparable with allowing that it can be said of an object *a* that it occurs equally in the *senses* of 'fa' and 'ga' (cf. 4.1211). To allow this would be to allow that a logical property of *a* can be described; but no such thing is allowed when we allow that the *fact* that *a* is combined with *f* in the same way as *a* is combined with *g* (if it is a fact) can be described. This must be a fact is *a*'s combining with *f* is a fact in the first place, just as *a*'s occurring in the senses of both 'fa' and 'ga' cannot be a fact if *a*'s occurring in the sense of 'fa' is not itself a fact. It is necessarily only internal or logical features of objects which are shown and unsayable, but whether *a* is actually configured with *b* in the same way as (say) *c* is configured with *d* is clearly not an internal or logical feature of these objects, and can be said.

RIGIDITY AND IDENTITY ACROSS POSSIBLE WORLDS

By M. J. MORE

SOME years ago Kripke introduced the now familiar distinction between rigid and non-rigid designation. He employed two criteria for rigidity:

(A) 'E' is rigid if 'E might not have been E' expresses a falsehood, i.e., if *being E* is an essential property of E;[1]

(B) 'E' is rigid if it designates the same thing in every possible world in which it designates anything and means E.[2]

But (A) and (B) are not equivalent. So I shall argue. To be sure, it seems that if 'E' is rigid under (A) then it is rigid under (B). But the converse does *not* hold. Kripke ought to have seen this. For an illustration of their non-equivalence is afforded by one of his own examples.

Consider 'the father of Elizabeth II'. According to Kripke[3], Elizabeth II could not but have had the parents she had: she could not have been born to different parents. On this view, then, no-one other than George VI could be the father of Elizabeth II. This is equivalent to 'the father of Elizabeth II' designating George VI in every possible world in which it designates something and means *the father of Elizabeth II*. But it obviously does *not* follow that it is not possible for George VI not to have been the father of Elizabeth II. He might have remained celibate all his life. Or he might have fathered Princess Margaret and no-one else. If so, he would not have been the father of Elizabeth II. Hence, 'the father of Elizabeth II' is rigid under (B), but is not rigid under (A).

The lack of equivalence of (A) and (B) arises in cases where *a* could not but have stood in the relation R to *b* (e.g. Elizabeth II could not but have been fathered by George VI), but *b* need not have had the property of standing in the converse of R to *a* (George VI need not have fathered Elizabeth). What such cases reveal is the possibility of something's being such that nothing other than it could have a certain property, e.g., *being father of Elizabeth II*, without it being the case that it could not but have had that property.

It may, of course, be contended that Kripke is wrong to espouse such a view concerning one's parentage. But even if this view is discarded, there are other, less controversial, examples which conform to the same pattern.

---

[1] See Kripke's 'intuitive test' for rigidity, 'Identity and Necessity' in *Identity and Individuation*, ed. M. K. Munitz (New York U.P. 1971), pp. 148–9; and footnote 25 to his *Naming and Necessity*, (Oxford 1980).

[2] 'Identity and Necessity', p. 145: 'What do I mean by "rigid designator"? I mean a term that designates the same object in all possible worlds.'

[3] *Naming and Necessity*, pp. 111–13.

Consider the pair set {Julius Caesar, Mark Antony}. Call it 'Caesant'. Then, given the criteria of identity for sets, nothing other than Julius Caesar could be the member of Caesant not identical with Mark Antony, i.e., 'the member of Caesant not identical with Mark Antony' designates Julius Caesar in every possible world in which it has the same meaning and designates anything at all. So it is rigid under (B). But it's possible for Julius Caesar not to have had the property of being the member of Caesant not identical with Mark Antony. Had Mark Antony not existed then neither would the pair set comprising him and Caesar have existed. And so, since it's possible for Caesar to have existed without Antony existing, it's possible for Caesar to exist without being a member of Caesant. Hence, it's not an essential property of Caesar (i.e., a property of Caesar whenever he exists) that he be the member of Caesant not identical with Mark Antony. Thus 'the member of Caesant not identical with Mark Antony' is rigid under (B) but not under (A).

Perhaps none of this is terribly important. But it seemed worthwhile making the point given the popularity of Kripke's distinction. It seems worthwhile, also, to note that 'Nothing other than x could be F' does not entail 'x could not but have been F'; that is, there are properties which, if not essential, aren't quite accidental either.

University of Hull

© M. J. More 1982

INTENTIONALITY AND INTUITION: A REPLY TO DAVIES

By E. J. LOWE

IN this paper I want to make a brief reply to Kim Davies's article 'Killing people intentionally, by chance' (ANALYSIS 41.3, June 1981), in which he criticises my earlier paper 'Peacocke and Kraemer on Butler's Problem' (ANALYSIS 40.3, June 1980). (The problem was originally presented by R. J. Butler in ANALYSIS 37.3, March 1977.)

It seems clear from the way that Butler presents his problem that *he at least* thinks that what one would intuitively say is intentional of Brown in the gun case is that he brought about Smith's dying *on this particular occasion*, i.e., as a result of this particular aiming of the gun and squeezing of the trigger. For surely it *is* clear that what he thinks one would intuitively say is *not* intentional of Brown in the dice case is that he brought about the die's falling with the six uppermost *on this particular occasion*, as a result of this particular shake and throw. Evidently, if Butler was *not* concerned, in *both* cases, with the intentionality or non-intentionality of Brown's doing something *on a particular occasion* (rather than, say, *sooner or later*, as a result of any one of a number of different attempts), the whole "problem" was bogus from the start, because it must then have rested purely on an equivocation.

However, Davies apparently takes the view (which I fully share and was at pains to defend) that it would be *mistaken* to say of Brown in the gun case that he intentionally brought about Smith's dying on this particular occasion (thus 'we ... have to say that it was not intentional that *on this particular occasion* Brown kill Smith' (Davies's emphasis) and, again, '[w]hat we cannot allow ... is to say that it is intentional that on this particular occasion ... Brown kill Smith'). But then it is not at all clear to me wherein he and I differ, or why he should dispute my suggestion that a careful examination of the two cases reveals an inconsistency in our intuitive ascriptions of intentionality. I can only suppose that he does *not* share what I take to be Butler's view, namely, that one would intuitively say of Brown in the gun case that he intentionally brought about Smith's dying *on this particular occasion* (a view, incidentally, which Ross, Kraemer and Peacocke, in their various contributions to the discussion, not only clearly *do* share but actually contend—fallaciously, as I believe—does *not* imply any inconsistency in our intuitions). If so, this only goes to show that Davies's intuitions in such matters are rather more logical than those of some of the rest of us. But *de intuitionibus non est disputandum*.

University of Durham

© E. J. LOWE 1982

MY BODY AND I—A REPLY TO FAHRNKOPF

By MICHAEL WELBOURNE

ROBERT FAHRNKOPF ('Descartes, Welbourne, and Indubitable Beliefs', ANALYSIS 41.3, June 1981) has misunderstood what I was up to in 'Cartesian Madness' (ANALYSIS 40.1, January 1980). I advanced no general view about the aims of the *First Meditation* but if I had it would have been much the same as Fahrnkopf's. In particular, I do not think there is a single belief mentioned in that Meditation which escapes Descartes's doubt. To be sure some beliefs are more resistant to doubt than others, but to be resistant is not to be impregnable. Descartes can use his dreaming-strategy to undermine all his beliefs as to what exists—at least until he lights on the belief that he exists. Among the beliefs which fall is the belief that there is a body answering, so to speak, to his 'bodily' sensations. A man who doubts whether there is such a body to be his could not rationally believe: '*This* body is mine'. *Pace* Fahrnkopf, this was one of *my* points—well taken it would seem.

Fahrnkopf may have been misled by my remark that 'Descartes *never* deviates from the position that it would be madness for me to doubt *concerning* these hands and this body (= mine) that they are mine'. But this does not mean that as long as I am sane I am bound to *believe*: 'These hands and this body are mine'. The claim I made, on Descartes's behalf, was, in effect, that there are certain forms which rational doubt cannot take, not that there are certain beliefs which a rational being cannot call into question. Even the dreaming-strategy will not enable me to doubt *concerning* this body (= mine) that it is mine: but in the rational pursuit of uncertainty there will be no need for me to doubt in that way if there is a reason to doubt whether there exists a body at all, to be mine.

My intention was to clarify just one step in the process whereby Descartes develops his doubts, a step which is completely obscured by the translation of Anscombe and Geach and necessarily overlooked by Fahrnkopf to the extent that his purposes are 'equally well served by either translation' (Fahrnkopf p. 139). The point is of wider interest, however, because this is the first appearance of an important *leitmotiv*. It certainly occurs again, much more prominently, in the *Sixth Meditation*, most obviously in the passage, quoted by Fahrnkopf, where Descartes speaks of the body 'which by a certain particular privilege I called mine'. Here he is commenting on his 'pre-critical' beliefs and in particular on the belief that this body 'belonged to me more properly and strictly ('magis ad me pertinere') than any other' (Lafleur translation 130). The passage to which I appealed—'the sailor in the ship'[1]—comes from his considered 'post-critical' discussion.

On this there seems to be a real philosophical *cum* exegetical disagreement between me and Fahrnkopf. According to him, the point of this passage is to show that 'one's mind and one's body are more intimately related to one another than the pilot is to his ship'. In particular, it does not show that 'Descartes takes Welbourne's position, or any particular position at all, on the question of how body-ownership is determined' (Fahrnkopf p. 139).

I am inclined to agree that this passage says nothing about body-*ownership*; neither did I. There is really no more reason to suppose that I *own* this body by virtue of the fact that it is mine than that I *own*, say, my view about Descartes's reference to madness. The relationships which may be referred to by uses of 'my', 'have', the genitive case etc. are immensely varied. Consider, for example, 'my Leibniz', 'my theory', 'my department', 'my accent', 'my umbrella'. Here it is only my umbrella which is, in any clear sense, owned; and yet even here it could be a question whether I own my umbrella, as it is a question whether I own my house. The problem is, what *is* the relationship which makes it proper for me to speak of this body as *mine*? Fahrnkopf may be prevented from seeing that this question is in the air in 'the sailor in the ship' by his insistence that it is concerned with the relationship between one's mind and one's body. But in fact Descartes does not so much as mention *mind* in the substance of the passage, but only in a theoretical postscript. The substance is about the relationship between *me* and my body—i.e. that relationship, whatever it may be, to which I refer when I speak of *my* body. Interestingly enough, Descartes (but not his French translator) drops the possessive adjective from *body* when he describes how it would be if I were merely present in a body as a sailor is present in a ship; it is as if he thought that *that* body would not be *my* body.

'The sailor in the ship' is difficult and not to be cited lightly. The fault is mine. But, if I am right about the madness passage, we may be able to use *it* to shed a little light on this later text. There we may learn at least that my body and I are so related that I should have to be mad to think: 'This body (= mine, the one of whose states I am often apprised through bodily sensations) is not mine'.

How is that relationship constituted? This is the difficult theoretical question. Descartes thinks—his theory requires him to think—that the relationship obtains by virtue of the mind and the body being related in a particular way, so as to give rise to a certain class of *confused* thoughts—specifically, thoughts, sensations, in having which I am immediately acquainted with states of my body. We can perhaps conceive that I should be aware of those same states in a different way. I should have to be if I were only related to a body as a sailor is to the ship on which he is borne along. In that case, being nothing other than a thinking thing, I should have to apprehend damage to the body by an intellectual analogue of the vision by which the sailor is apprised of damage to the ship. But manifestly this is not how things are in fact. I am so related to my body that I have an intimate 'inward' way of apprehending my injuries and bodily needs—in having sensations of pain, hunger, thirst etc. The 'particular privilege' by which I call this body mine derives from this relationship.

---

[1] In 'Cartesian Madness' I wrongly used the more memorable, but misleading phrase—'the pilot in the ship'—which we owe to Descartes's French translator. Descartes's *nauta* may not be at the helm; nor does the sense of the passage require us to suppose that he is.

Descartes tells us that these modes of consciousness arise from the 'permixture as it were' ('quasi permixtio') of the mind with the body. This seems to be a metaphor ('quasi'), drawn perhaps from seventeenth century science. The great problem is how to cash it. Descartes is no help. He remarks to Burman that these matters are 'very difficult to explain; but here our experience is sufficient, since it is so clear on this point that it just cannot be gainsaid' (*Descartes' Conversation with Burman, tr.* John Cottingham, p. 28). Perhaps we may infer from Descartes's resort to metaphor that at this crucial point the phenomena were just too much for his theoretical ingenuity.

University of Bristol © MICHAEL WELBOURNE 1982

ON THE TRANSITIVITY OF EPISTEMIC PREFERABILITY

By P. J. R. MILLICAN

R. A. SORENSEN has argued recently that the 'more likely to be correct than' relation is not transitive ('Is Epistemic Preferability Transitive?', ANALYSIS 41.3, June 1981). This result is certainly surprising and would, if correct, rule out any treatment of epistemic preferability which conformed to the calculus of probabilities, since any quantitative measure of probability, being numerical, must necessarily give rise to a relation of 'having greater probability than' which *is* transitive. Sorensen's argument, however, purports to demonstrate the non-transitivity of epistemic preferability without at any point assuming that epistemic preferability is non-probabilistic. This should make us very suspicious.

Let us imagine a game involving three dice, each with six faces which are numbered as follows:

Die A: 1, 2, 9, 14, 15, 16
Die B: 6, 7, 8, 11, 12, 13
Die C: 3, 4, 5, 10, 17, 18

The two players each choose one of these dice, and they throw their chosen die simultaneously. The winner is the player whose die shows the higher score, and we can suppose that the loser is then obliged to pay him one penny.

The interesting thing about this game is that A will score more than B in 21 of the 36 possible combinations (giving a 7/12 chance of A's winning over B, since all combinations are equally probable), while B will in turn score more than C 21 times out of 36 and, surprisingly, C will also defeat A in the same proportion. Thus a gambler who allows his opponent the first choice of a die can be fairly sure of making a profit by choosing his own die appropriately—for every 12 'rounds' played he can expect to gain two pence, by winning 7 rounds and losing 5.

Sorensen claims that this strange phenomenon can provide a counterexample to the transitivity of epistemic preferability. For instance, suppose that three research groups are each given a throw of dice A, B and C respectively, and that money is allocated to each research budget in proportion to the score achieved—say £7,000 for a score of 7; £10,000 for a score of 10 etc. If we assume that the three research groups are attempting to solve three equally difficult problems, and that a greater budget brings a greater chance of success, then, since it is likely that the first group has a larger budget than the second, Sorensen would have us conclude that its results are epistemically preferable to the results of the second group. Similar reasoning would conclude that the results of the second group are epistemically preferable to the results of the third, and the results of the third to those of the first. Epistemic preferability, then, cannot be transitive.

As I suggested at the beginning of this article, the flaw in Sorensen's argument can be seen most clearly if we assume (for the benefit of his attempted *reductio ad absurdum*, if you like) that epistemic preferability *is* probabilistic. Let us suppose that a research group's chance of success is directly proportional to its budget, for example, a budget of £7,000 giving a 7% chance of correct results, etc. Now, it is true that the budgets of the first two groups are *probably such as* to give the first group a greater *chance* of success than the second; similarly with groups two and three, and with groups three and one. But it obviously *cannot* be right to infer the absurd conclusion that the results of group one are more probable than the results of group two, which are more probable than the results of group three, which are more probable than the results of group one—'mathematically more probable than' *must* be transitive!

Sorensen's mistake is to assume that, because the first group's results are *probably more probable* than the results of the second group, they are therefore more probable *tout court*. In fact, since the average of the numbers on each of the dice A, B and C is 9½, the average expected budget of each group turns out to be £9,500, and the chance of success 9½%. The three groups are equally likely to produce correct results, and so there is no epistemic preference amongst their conclusions.

I shall conclude with another, more dramatic, illustration of Sorensen's error. Imagine that a criminal, tied to a post, has the choice of two alternative punishments. If he chooses the first, he will be 'shot' once with a catapult; if he chooses the second, he will be shot once with a revolver whose barrel, containing two bullets and four blanks, has already been spun randomly in front of his eyes (as in Russian roulette). Now, it is *probable* that the barrel is positioned so that the first punishment is more *likely* than the second to inflict a serious injury—in four chances out of six (corresponding to possible positions of the barrel in which a blank, rather than a bullet, is in the firing position) it will be quite impossible for the second punishment to inflict any injury at all. Nevertheless, the rational criminal will obviously choose the first punishment in preference to the second. A high probability of serious injury in two chances out of six outweighs a very small probability in six out of six.

Lincoln College, Oxford © PETER MILLICAN 1982

CONSCIOUS BELIEF

By N. M. L. NATHAN

HOW is it', asks D. H. Mellor, 'that I can unhesitatingly answer almost any "yes or no" question I understand? My answer may be "I don't know", but how do I know that? . . . My knowledge of my beliefs and doubts, though fallible, is vast, and immediately available to me as it is to no-one else. Why is that?' ('Conscious Belief', *Proceedings of the Aristotelian Society* 78, 1977–8, p. 87). The short answer, according to Mellor, is that 'when a question is put to me I become conscious of my belief or doubt on the matter'. This is, he thinks, a true answer, so far as it goes. 'The experience of conscious belief or doubt is familiar enough, as is the fact that it is what enables us to answer questions, and generally to converse.' But Mellor also thinks we need a clearer understanding of conscious belief itself, and goes on to argue that for me to be conscious of my belief that p is simply for me to believe that I believe that p, where belief itself is to be understood according to what he calls the 'action theory.' According to this belief is a mental state which, though not a disposition to action, is inferred in the explanation of action: actions are caused by conjunctions of beliefs and desires. Believing that one believes, as opposed to plain believing, is required for the explanation of linguistic action, and itself requires the existence of an inner sense. (His theory is further elaborated in 'Consciousness and Degrees of Belief', D. H. Mellor (ed.), *Prospects for Pragmatism* (Cambridge 1980) pp. 139-173).

One objection to Mellor's theory is that it seems possible for me to believe that I believe p, as this would be explained by the action theory, even while my attention is completely absorbed by matters which have nothing to do with myself or my own mental states or, indeed, p. Imagine that I am in a state whose conjunction with my wanting someone to believe the truth about me would explain my saying 'Yes' to him were he to ask me whether I believed that $7 + 5 = 12$. On the action theory this state is of believing that I believe that $7 + 5 = 12$. But I can be in this state even if the man has not in fact asked me this question, and at a time when the only thing I am conscious of is, say, intense cold. Mellor would no doubt reply that although I can believe that I believe that $7 + 5 = 12$ when the only thing I feel is intense cold, it does not follow that I can believe that I believe that $7 + 5 = 12$ when the only thing I am *conscious* of is intense cold. This does not follow because consciousness of having a belief does not require you to have feelings or sensations. Consciousness of having a belief needs clarification precisely because it differs in this way from other states of consciousness like feeling cold or having a visual sensation. How can it be clarified if it is not identified with believing that one believes? Against this, I would maintain that there is something or other it feels like to be conscious of anything, and that it is a mistake to suppose that linguistic action can be explained in terms of believing that one believes only when the agent has the feelings inseparable from consciousness of his belief. Mellor approvingly quotes Kneale: 'When we realise that $2 + 2 = 4$ we do not sweat with any feeling of supreme intensity'. And no doubt the strength of a belief of which I am conscious is not measured by the 'strength' or 'vivacity' of what it feels like to be conscious of it. But it does not follow that we do not have any feeling at all when we attend to the proposition that $2 + 2 = 4$ or are conscious of believing it. Nor is it even clear that Mellor's own theory does not commit him to saying that consciousness of believing requires sensation: it is after all an inner *sense* that he postulates in order to explain our knowledge of what we believe.

But my main objection to Mellor's account of what it is to be conscious of believing p is that he obscures the existence of a phenomenon which allows us to give a different and perhaps better explanation of why we have such vast and immediately available knowledge of our beliefs and doubts. As a label for conscious belief, in the sense of consciousness of one's believing, Mellor appropriates, with some misgivings, the term 'assent'. To assent to p, on his usage, is to be conscious of believing p. Now, as traditionally used, 'assent' stands for a decision about a proposition, sometimes and perhaps misleadingly called an act of belief, but at any rate productive of a state of belief. The phenomenon to be captured by this talk of assenting to p was, I suggest, the decision to act as if p is true and not to act as if p is false. If you assent to p, in this sense, your assent is conscious, not necessarily in the sense that you are conscious of it, but certainly in the sense that it requires consciousness on your part of what you assent to. If this assent is called an act of belief, there is a sense of 'conscious belief' in which it stands not for states of belief of which the owner is conscious, but actions of believing in which the agent is conscious of the proposition believed. But however this may be, traditional assent at least causes (produces or sustains) a state of belief. And this fact allows for a simple explanation of how it is that I have such vast and immediately available knowledge of my beliefs and doubts. If you ask me whether or not I believe p, you give me the opportunity, by assenting to p, to bring about in myself or sustain the state of believing that p. If I assent to p, and know that I have done so, I know that I shall believe p at the time I say to you that I do. Instead of having introspective knowledge of a belief-state whose conjunction with desire causes action, I have introspective knowledge of a decision, and inferential knowledge that this decision causes a belief-state whose conjunction with desire causes action. We can also explain how it is that when you are conscious of a proposition which you believe, you do sometimes at least have feelings unlike those you have when you merely consider or entertain the proposition. You have these special feelings when you are conscious of the assenting which causes the belief-state and what you feel is what it is like deciding to act as if the proposition you assent to is true.

Why should your deciding to act as if p is true and not to act as if p is false cause a corresponding belief-state, given that it is perfectly natural to say: 'I shall act as if p is true but in fact I am not at all sure that it is'? I suggest that, taken in context, the first part of such a remark means 'I shall act for such and such particular purposes as if p is true'. I might, for example, be deciding to act as if p is true just in my capacity as nightwatchman or defending counsel. Assenting to p, however, is deciding to act for all purposes as if p is true. If I assent to p and then say with sincerity that I am not sure that p is true, that speech act is an action of precisely the sort that, in assenting to p, I am deciding not to perform.

Finally, there is a problem about whether one can describe the content of a man's decision to act as if p were true without using the notion of consciousness of belief. Suppose I decide to act as if Arthur is innocent. Do I decide to do things like defending his reputation when other people attack it, or things like defending his reputation when I believe that other people are attacking it? You might think that if my decision is to produce action on my part it will have to be a decision of the latter kind. For how can I know when to defend him unless I believe he is being attacked, and indeed unless I am conscious of believing this? But then how can you understand assent, in the traditional sense, without understanding what it is for a man to be conscious of what he believes, and how can an explanation, in terms of assent, of the vastness and immediate availability of a man's knowledge of his beliefs, be simpler than an explanation in terms of his introspection of his belief-states? My answer to this objection is that a decision does not need to be capable of being carried out in order for one to give an accurate description of its content, and that actions can be influenced by decisions incapable of being carried out. Just as the accurate description of a belief can give it a content which determines no definite proposition, so, on an accurate description of its content, a decision can be impossible to implement except by accident. I cannot sometimes know that now is the time to defend Arthur's reputation unless I sometimes believe that now it is being attacked. I may even need to be conscious of having such a belief. But this does not stop me from being caused to defend his reputation on believing it is attacked by my decision simply to defend his reputation when it is attacked.

University of East Anglia © N. M. L. NATHAN 1982

THE DISPOSITIONAL/CATEGORICAL DISTINCTION

By ELIZABETH W. PRIOR

THE commonly accepted view is that dispositions differ from categorical properties because the former possess a special relationship to subjunctive conditionals not possessed by the latter. More specifically one common view is that sentences which ascribe dispositions entail certain subjunctive conditionals while those which ascribe categorical properties do not.[1]

In his paper 'In Defense of Dispositions' D. H. Mellor argues that dispositions do not peculiarly entail subjunctive conditionals. (*The Philosophical Review*, 83, 1974, p. 173). More specifically, Mellor argues that the *paradigms* of categorical properties used by Categorical Realists to solve the "problem" of dispositions in fact *share* with those dispositional properties the feature of entailing certain subjunctive conditionals.

> Take the paradigm, molecular structure—a geometrical (for example triangular) array of inertial masses. To be triangular is at least to be such that if the corners were (correctly) counted the result would be three. (p. 171)

So, according to Mellor, while it may be the case that dispositions must have bases these bases need only to be other dispositions. And he goes on to say:

> No doubt there are virtues in explaining properties of things in terms of other properties especially in terms of those of their spatial parts . . . But none of this gives reason to suppose that dispositions can or need to be made respectable by being given some non-dispositional basis. (p. 172)

Mellor's claim is, in effect, that while statements attributing dispositions to items do entail certain subjunctive conditionals, so do categorical ascription statements. According to Mellor dispositional and categorical ascription statements are alike in that both entail subjunctive conditionals. Look again at what he says about triangularity (p. 171). This amounts to

(1) 'x is triangular' entails 'If the corners of x were (correctly) counted the result would be three.'

There are two ways in which one may argue against Mellor and for the thesis that dispositional ascription statements possess a relationship to certain subjunctive conditionals not possessed by categorical ascription statements. The first way turns on the difference between (1) and (2):

(2) 'x is triangular' entails 'If the corners of x were counted the result would be three.'

(2) is a closer analogue than (1) of the (clearly) dispositional (3):

(3) 'x is fragile' entails 'If x were (suitably) dropped x would break.'

Now (2) is false. For someone (in this world) may count badly and get the wrong answer. More importantly there is another possible world in which the laws of nature are different and a *systematic* perceptual deception occurs—instead of seeing and counting three corners the inhabitants see and count five corners whenever confronted by a triangle and the order to count.

(1) differs from (2) in containing the word 'correctly'. Since (2) is clearly false, we may conclude that Mellor sees the insertion of the word 'correctly' as securing the entailment in (1). How then ought we interpret 'correctly'?

In a footnote (fn. 35) to the passage quoted from page 171, Mellor says

> 'Correctly' here refers to how the counting is done not to whether it gives the result three.

---

[1] See Goodman, N., *Fact, Fiction and Forecast*, Bobbs-Merrill Co. Inc., 1955, p. 34. Goodman claims that this view is the commonly accepted one.

But if this is so then the addition of 'correctly' to the right hand side of (2) will serve no useful purpose. If 'correctly' is cashed in the terms Mellor suggests in footnote 35, (1) is false. It is false for the second reason given for claiming that (2) is false. (1) is false because there is another possible world where the laws of nature are different and a *systematic* perceptual deception occurs—instead of seeing and counting three corners the inhabitants see and count five corners whenever confronted by a triangle and the order to count. More generally, the point is that *manner* of counting cannot guarantee getting the right answer.

Unless 'correctly' is explicated in a way other than that suggested in footnote 35, (1) will be false. To make it even *initially plausible* that the entailment from left to right holds in (1) 'correctly counted' will (contra Mellor) have to refer to the *result* rather than to the method of counting. 'Correctly counted' will need to mean 'if the number *is* three then the result of the counting will be three, if the number *is* four then the result of the counting will be four etc.' I say initially plausible because I will go on to argue that even if we take 'correctly counted' to refer to the result, it still involves an ambiguity. Reading the ambiguity one way (1) is false, reading it the other way (1) is true, but the distinction between categorical and dispositional can still be made. Let us *suppose* for the moment however that the entailment does hold. But now (1) involves a kind of circularity. To be triangular is (at least) to *have* three corners, but the only possibly adequate cashing out of 'correctly counted' is the one just stated. In the case of genuine dispositions on the other hand we do not have such circularity. Take (3). Here we do have a conception of being suitably dropped that can be explicated *without* reference to the notion of fragility. I am not saying that 'being suitably dropped' is easy to explicate but I am saying that it is plausible that it can be explicated without reference back to fragility. This is precisely what *cannot* be done with 'being correctly counted' in the case of triangularity.

Now for the second way in which one may argue against Mellor. I have already argued that the only possible explication suitable for his purposes of 'correctly counted' is one which refers to the result and that this is open to objection. But even given we assume this there is an ambiguity to be resolved. The antecedent of the right-hand side of (1) with 'correctly' read in terms of result is ambiguous. On one reading of the ambiguity the entailment does *not* hold. On the other reading, though the entailment holds, we can still distinguish between dispositional and categorical properties in terms of relationships to certain subjunctive conditionals.

The right side of (1) can be read as either

(4) If one were to count the corners of x and get the *actually* correct answer the answer would be three.

(5) If one were to count the corners of x and get the *counter-factually* correct answer then the answer would be three.

Now the left-hand side of (1) does not entail (5) because there are worlds with different laws of nature where the counting of the corners *changes* the number of corners. Within *this* world there are similar cases. In the case of certain elementary particles, measuring the position of such a particle *changes* the position of that particle. Hence resolving the ambiguity in (1) the second way makes (1) false.

Reading the right-hand side of (1) as (4) we *do* get an entailment, (1) is made true. But the antecedent of (4) involves a reference back to the *actual world*. This stands in contrast to the antecedent of the conditional entailed by a truly dispositional ascription statement—the antecedent of the right-hand side of (3), for example—where there is *no* such reference back to the actual world.

Thus, the only reading of the right-hand side of (1) which preserves the entailment is one which still allows us to make the dispositional/categorical distinction.

I therefore conclude that the commonly accepted view that dispositional properties can be distinguished from categorical ones because dispositional ascription sentences possess a relationship to certain subjunctive conditionals not possessed by categorical ascription sentences survives Mellor's attack unscathed.[2]

Monash University

© ELIZABETH W. PRIOR 1982

COUNTING CORNERS CORRECTLY

By D. H. MELLOR

ELIZABETH PRIOR tries to sustain the myth that there are non-dispositional properties of things, i.e. properties that support no subjunctive conditionals, by denying my claim that 'x is triangular' entails 'If x's corners were correctly counted the result would be three', where 'correctly' refers (i) to how the counting is done, not (ii) to whether it gives the result three. She does not succeed.

I made my stipulation (ii) about 'correctly' to prevent my claim being trivially analytic, and specifically to prevent a counting counting as correct which got the right answer by compensating errors: e.g. counting one corner twice and missing another altogether. As to (i) I thought what I meant was too obvious to need spelling out, but apparently not. So:

---

[2] I wish to thank Frank Jackson and Robert Pargetter for criticism and advice.

to count things correctly, count them all once each, i.e. put them all in some 1–1 correspondence with an initial segment of the sequence of positive integers 1, 2, 3, ... The highest number in the segment is the result of the counting.

That is how to count correctly any collection of things, and if the corners of any triangle were so counted, the result would, as I claimed, be three. *Pace* Prior, there is no circularity in this claim, unless all obviously necessary truths are to be dismissed as circular. Maybe, as Prior suggests, some collections of corners (or other things) cannot be counted, either at all or without changing their number, which must therefore be found out by means other than counting. But that is no objection to my claim. In the article she refers to, I went to some pains to point out that subjunctive conditionals do not entail the possibility of their antecedents, and are therefore not contradicted by their antecedents' impossibility.

Cambridge University © D. H. MELLOR 1982

VANISHING POWERS: A REPLY TO MILLER AND WILSON

By HILLEL STEINER

DAVID MILLER (ANALYSIS 41.4, October 1981, pp. 222–3) and Patrick Wilson (ANALYSIS ib. pp. 219–21) reject my criticism of Nozick's attack on Hart (ANALYSIS 41.1, January 1981, p. 50). I'm afraid I remain unconvinced by their attempts to rescue Nozick from my charge of self-contradiction. In what follows, I hope the reasons for that charge emerge more clearly.

Nozick's view is that a right to something does not, contrary to Hart, imply a right to enforce that right. Nozick is also of the view, shared by Hart, that to be a promisee is to be in possession of a right to what was promised. Hence he suggests that there is nothing logically objectionable about (i) X, by virtue of his promise P_1, conferring a right upon Y to X's doing A, *and* (ii) Y, by virtue of his promise P_2, conferring a right upon X to Y's non-enforcement of X's doing A.

It follows from Nozick's thesis that there can similarly be nothing logically objectionable about (iii) X, by virtue of his promise P_3, conferring a right upon Y to X's non-enforcement of Y's non-enforcement of X's doing A, nor about (iv) Y, by virtue of his promise P_4, conferring a right upon X to Y's non-enforcement of X's non-enforcement of Y's non-enforcement of X's doing A.

Unfortunately, X's initial shrewdness fails him toward the end of this bargaining process and he consequently neglects to extract P_4 from Y as part of the exchange of promises. Thus although Y has an obligation not to enforce X's doing A, he nevertheless remains at liberty—has no obligation not—to enforce X's non-enforcement of Y's non-enforcement of X's doing A. In the event, X's oversight costs him dearly for, when he attempts to default on his obligation to do A, Y forces him to do it. Shameless to the end, X denounces Y and, invoking Nozick's anti-Hart thesis, protests that not only does Y's right to X's doing A fail to imply Y's right to enforce X's doing A, but also that (under P_2) Y is actually obligated not to enforce it.

The ingenuous Y is taken aback and, anxious to prove himself an equally loyal adherent of the Nozickian thesis, pleads innocent to the charge of having broken his P_2 promise. 'All I did', he insists, 'was to enforce your non-enforcement of my non-enforcement of your doing A. We both agree with Nozick that rights—including rights to enforce rights—are severally created by separate promises, and that having a right to something in no way implies a further right to enforce that right. I gave you no P_4 promise and what I did was therefore quite permissible.'

X's opportunistic rejoinder is quick and to the point. 'I extracted no P_4 promise from you because I presumed—and events have borne me out—that the obligation you would have incurred under P_4 was one you had already incurred under P_2.' 'Not quite', replies Y. 'It's certainly true that all P_2 forbearances are P_4 forbearances. But the reverse is not true, for there are measures I could have taken to force you not to force me not to force you to do A, without actually forcing you to do A. Since P_4 is *necessarily* extensionally non-equivalent to P_2, there are no grounds for your presumption that the right conferred upon you by the latter logically implies the right that would have been conferred upon you by the former. Let me hasten to add that I do understand why you think I was obligated not to act as I did. But you have given me no reason why I should regard the obligatoriness of P_2 forbearances as overriding the non-obligatoriness of P_4 forbearances. The charm of the Nozickian thesis precisely lies in the fact that it allows my enforcing action to be judged both impermissible and permissible.'

University of Manchester © HILLEL STEINER 1982

ROWE, SELF-EXISTENCE, AND THE COSMOLOGICAL ARGUMENT

By B. F. KEATING

WILLIAM L. ROWE'S introductory text, *Philosophy of Religion*, contains an interesting discussion of that part of the Cosmological Argument which is supposed to establish the existence of a self-existent being (Encino: Dickinson, 1978, Chapter Two). A self-existent being is 'a being whose existence is accounted for by its own nature' (p. 18). In contrast, a dependent being is 'a being whose existence is accounted for by the causal activity of other things' (p. 18). The relevant part of the Cosmological Argument is set out by Rowe as follows:

1. Every being (that exists or ever did exist) is either a dependent being or a self-existent being.
2. Not every being can be a dependent being.

Therefore,

3. There exists a self-existent being (pp. 18–19).

Of this argument Rowe says, 'its premises imply its conclusion and its premises do seem to follow from PSR' (p. 29). PSR is a version of the Principle of Sufficient Reason. It states that there must be an explanation of any positive fact whatever (p. 21). Does Rowe mean that we must accept the existence of a self-existent being? No, for

> Since we do not seem to know that PSR is true we cannot reasonably claim to know that the premises of the Cosmological Argument are true. They might be true. But unless we do know them to be true they cannot *establish* for us the conclusion that there exists a [self-existent] being. . . . If it were shown, however, that even though we do not *know* that PSR is true we all, nevertheless, *presuppose* PSR to be true, then, whether PSR is true or not, to be consistent we should accept the Cosmological Argument (p. 29).

I believe that this representation of the Cosmological Argument is incorrect. In the first place, I shall show that it is not true that the premises follow from PSR. It must be premised in addition that it is logically possible that there exists a self-existent being. Yet Rowe does not show, or argue, that this is possible. In the second place, I shall show that there is a reasonable interpretation of 'self-existent' such that, if it is *possible* that a self-existent being exists, then it is *actually true* that one exists, independently of whether or not PSR is true. Hence, if this interpretation is accepted, then, insofar as the concern is to establish that there exists a self-existent being, PSR is (I) insufficient if the notion that such a being exists is not coherent, and (II) unnecessary if that notion is coherent. And it is very difficult to see how 'self-existent' is to be interpreted if my interpretation of it is rejected.

I

Let us begin by considering Rowe's claim that premise 2 follows from PSR. Rowe argues:

> For if the question to be answered is why there are (or have ever been) any dependent beings at all, we cannot answer that question by noting that there always have been dependent beings, each one accounting for the existence of some other dependent being. Thus on the supposition that every being is dependent it seems there will be no explanation of why there are dependent beings (p. 27).

However, this last claim does not follow. For it may be that, though we cannot explain dependent beings simply by noting that there always have been dependent beings, we nevertheless can explain dependent beings by appeal to the thesis that, as regards causal status, dependent beings are the only kind of beings there could be. And this thesis follows from PSR plus the claim that the notion of a self-existent being is incoherent. After all, it is logically impossible that, in general, the explanation of why there are concrete entities of kind K should *require* the introduction of concrete entities not of kind K. Where kind K is, for example, *being a cow or not a cow*, it is logically impossible that there should exist concrete entities not of kind K. Yet PSR could still be true for all that. Moreover, given the assumptions of the Cosmological Argument, it is impossible that explaining K's should require introducing non-K's, even if non-K's may be possible when considered apart from the Cosmological Argument's assumptions. Where kind K is, for example, *being something for which there is an explanation*, it is logically impossible that PSR is true and that there should exist entities not of kind K. Given that there are beings (and I here postpone the question of how the proponent of PSR might explain why there is something rather than nothing) and that PSR is true, the explanation of why there are *beings which are either cows or not cows*, or of why there are *beings which have explanations*, might simply be that there could not consistently be beings not of these kinds. Similarly, given that there are beings and that PSR is true, the explanation of why there are dependent beings could be that, from the point of view of causal status, there could not consistently be beings not of this kind. Thus, PSR does not imply premise 2, viz that not every being is dependent. It must be premised in addition that it is logically possible that there exists a being whose existence is accounted for by its own nature.

II

I now offer a reasonable interpretation of 'self-existent' such that, if it is possible that there exists a self-existent being, then there actually exists one. Suppose there is a possible world W in which X is self-existent. Then, (A) X's existence in W is uncaused, and (B) there is a "nature" (concept, form, property) N which *itself* accounts for X's existence in W. (A) follows since a self-existent being is not due to another's causal activity; no self-existent being is dependent. Nor can a self-existent being be explained by its own causal activity, for that presupposes its existence. (B) follows since it cannot be N's *instantiation*, either by X or by another, which accounts for X. Fires have natures (p. 11), and a fire's existence follows from the fact that *it* instantiates its nature, but a fire would presumably not be self-existent. And if X's existence was due to *another* N, X would be dependent on the other and not on "its own nature." At least, God, in being self-existent, is not supposed to be dependent on another. Now, the only sense I can make of the conjunction of (A) and (B) is by supposing that, in W, the existence of N entails the existence of X; i.e., that N is a definition of X in W. After all, if the connection between N and X which accounts for X is not causal, and not conceptual or logical, what is it? But if this is correct so far, it follows that X actually exists, if X is possible. For, if X is possible, then N is actually consistent. If N were not in fact consistent, then it would be impossible for X to exist since N is X's nature. But if N is actually consistent then N exists. Now we are supposing that, in W, N entails, or defines, X. But what a property or quality entails, or is a definition of, does not vary from world to world. So N entails, or is a definition of, X in every world. But we have already shown that N exists. So X exists. Thus, on this interpretation of 'self-existent,' if the notion is even coherent, if it is even possible that there exists such a being, then there actually exists such a being, independently of whether or not PSR is true.

This interpretation of 'self-existent' accords with Rowe's definition. Moreover, it confirms his commitment to the possibility of self-existents. For, granting a realism about abstract entities, it seems that we can produce examples of self-existents, albeit from non-theological contexts. Thus, the following would presumably qualify as natures: being the ruler and compass construction of a line perpendicular to a given line; being the triangle the square on the hypotenuse of which is equal to the squares on the other two sides. And the proof that these natures are consistent amounts to a proof that there actually exist such abstract entities. The construction and the triangle may be said to be self-existent in that their existence is not due to any causal activity but rather is due to (the consistency of) their own natures or definitions. Finally, our interpretation allows the theist to explain why there is something rather than nothing by saying that it is logically necessary that the self-existent (the definable) beings exist. If the ruler and compass construction does not exist in some world W, it must be because the property of being that construction is inconsistent in W. But that is absurd. The property is just the actual property we know to be consistent, and what is logically possible does not vary from world to world.

I have argued that PSR is insufficient to establish the existence of a self-existent being. If the above interpretation of 'self-existent' is accepted, then PSR is also unnecessary. However, if the existence of a self-existent being is *already granted*, then PSR may suffice to show that such a being is crucial to explaining the world. It is with regard to this point that I think Rowe's interpretation is strongest. For certain kinds K, we can explain why there are beings of kind K by supposing that a being not of kind K creates them, while we cannot explain this by supposing that another being of kind K creates them. But it does not follow that this holds for any kind K. It depends on whether or not the notion of a being not of kind K is coherent, and in the specific case of the notion of a self-existent being this coherence is controversial, as I am sure Rowe would acknowledge. But, then, even if we know that PSR is true, it does not follow that 'to be consistent we should accept the Cosmological Argument.'

West Virginia Wesleyan College © B. F. KEATING 1982

OMNIPOTENCE AND SIN

By W. R. CARTER

THOSE who argue that God exists often assume, and do not attempt to demonstrate, that 'the concept of God' is consistent.[1] I believe that this assumption should be questioned. In what follows I shall take the property of being God to be the property a thing has if and only if this thing fulfils a certain role, or occupies a certain office or position. This position—Pavel Tichy calls it 'the divine office'—has an *essence* in the sense that there are certain conditions, or 'requisites', that any individual occupant must satisfy.[2] Generally it cannot be assumed that the essence of an office is the same as the essence of the individual holding this office. (It is necessarily true that any individual who fulfils the role of being Edna's husband is married. So being married is part of the essence, as Tichy would say, of this role or office. Nonetheless, being married is not essential to Boris—the individual who happens to *be* Edna's husband. Boris *can* exist after ceasing to be married.) The exact requisites for holding the divine office are a matter of some dispute. Should these requisites be logically inconsistent—should they be such that it is logically impossible for *any* individual to satisfy them—it is necessarily false that God exists.

What are the requisites for occupying the divine office? Non-malevolence certainly deserves a place on a list of requisites. Tichy argues that no individual is such that it is essentially non-malevolent. (p. 411) If Tichy is right about this, and if non-malevolence is indeed a requisite for the divine office, it follows that no individual that occupies this office can be such that it is essentially an occupant of the office in question. If there is indeed an occupant of the divine office, this individual is such that it can (as we would normally say) *cease to be* God.

Like Tichy, Nelson Pike is committed to the thesis that no occupant of the Godly office or position can be such that it *must* occupy this office.[3] Briefly stated, Pike's position is this: both omnipotence and perfect goodness are requisites of the divine office (pp. 70 and 78). Letting 'D' designate this office or position, Pike endorses both:

$$(1)\ \mathrm{Nec}(x)(\mathrm{occupant}(x, D) \rightarrow \mathrm{omnipotent}(x)),$$

and

$$(2)\ \mathrm{Nec}(x)\ (\mathrm{occupant}(x, D)\text{— does not sin}(x)),$$

but also rejects

$$(3)\ \mathrm{Nec}(x)(\mathrm{occupant}(x, D) \rightarrow \text{cannot sin}(x)).$$

Pike's view is that the conjunction of (1) and (3) does, while the conjunction of (1) and (2) does not, pose inconsistent requisites for occupancy of D. In one place he observes that:

> The individual who bears the title 'God' (Yahweh) might have the creative power necessary to bring about objects or states of affairs the production of which would be morally reprehensible even though 'God sins' is logically contradictory. The conclusion is, simply, that if an individual bears the title 'God', He does not exercise this creative power. (p. 81)

And in another:

> The term 'God' has been so specified that the individual who is God *cannot* sin and be God. But it will not follow from this that the individual who is

---

[1] I have profited from reading James Cargile's paper 'The Ontological Argument,' *Philosophy* 50 (1975), pp. 69–79. Cargile argues that the 'question about the consistency of the concept of God is the fundamental question in the ontological argument.' (p. 69)

[2] Pavel Tichy, 'Existence and God,' *The Journal of Philosophy* Vol. LXXVI, No. 8 (August, 1979), p. 411.

[3] Nelson Pike, 'Omnipotence and God's Ability to Sin,' *American Philosophical Quarterly*, Vol. 6, No. 3 (July, 1969), pp. 208–216. This article is reprinted in *Divine Commands and Mortality*, ed. by Paul Helm (Oxford, 1981). The passage cited appears on page 68 of this anthology.

> God does not have the *ability* to sin. He might have the creative power necessary to bring about states of affairs the production of which would be morally reprehensible. He is perfectly good (and thus God) insofar as He does not exercise this power. (pp. 70–80)

What Pike does not explicitly say, but what is implied by what he does say, is that it cannot happen that an individual that occupies *D* is such that it must occupy *D*. Let us grant—for the sake of argument, if you will—that *D has* an occupant—one that is designated by the term 'Yahweh'. From the conjunction of (1) and

(4) $\text{Nec}(x)(x = \text{Yahweh} \rightarrow \text{occupant}(x, D))$.

it follows that

(5) $\text{Nec}(x)(x = \text{Yahweh} \rightarrow \text{omnipotent}(x))$.

(5) tells us that there are no possible circumstances in which Yahweh exists lacking the attribute of omnipotence. Yahweh is *essentially* omnipotent, if (1) and (4) are true. Moreover, since it follows from the conjunction of (2) and (4) that:

(6) $\text{Nec}(x)(x = \text{Yahweh} \rightarrow \text{does not sin}(x))$,

anyone who accepts the conjunction of (2) and (4) must accept the thesis that Yahweh is essentially a being that does not sin—is, in other words a being that *cannot* sin. This conclusion cannot be reconciled with Pike's claim that Yahweh, 'the individual who is God,' has the *power* to bring about states of affairs the production of which would be morally reprehensible. It is a corollary of (6) that there is no possible world in which Yahweh sins and so no possible world in which Yahweh brings about a morally reprehensible state of affairs. Since this conflicts with the assumption that Yahweh is omnipotent, it seems that Pike must reject (6). But (6) follows from the conjunction of (2) and (4). Since Pike accepts (2), he must reject (4). Thus Pike is committed to saying that the individual that occupies *D*, the divine office or position, is such that it is possible for this being to not be the occupant of this position.

Pike is correct in not endorsing the conjunction of (5) and (6). For this conjunction is inconsistent. More precisely, the hypothesis that there is such an individual as Yahweh, concerning whom (5) and (6) are true, is inconsistent. The world simply cannot be such that it contains a being that is both (essentially) omnipotent and essentially sinless. Any being that is essentially sinless is such that there is no *possible* world in which it commits sinful actions. Since such a being cannot sin, it cannot be all-powerful and so cannot satisfy one of the requisites of the divine office. Given the assumption that any being that occupies the divine office cannot exist without occupying this office, no being can satisfy the requisites of the divine office.

Since he believes the requisites for D are consistent, Pike rejects (4). But the Pike gambit is not successful. Any complete list of requisites for occupancy of office D surely must include:

(7) $\mathrm{Nec}(x)(\mathrm{occupant}(x, D) \rightarrow \mathrm{omniscient}(x))$.

Let us suppose—following the lead offered by Pike and Tichy—that Yahweh (a hypothetical occupant of D) is not *essentially* sinless, even though *in fact* Yahweh never sins. It is a corollary of this assumption that there are possible worlds in which Yahweh *does* sin. Let 'W_s' designate such a world and 't' designate a time, in the history of W_s, at which Yahweh acts in a sinful way. (If you deny that there is such a possible world as W_s then you are committed to saying that Yahweh is essentially sinless.) It follows from (2) that Yahweh does not occupy D after t in W_s. So (4) must be false, if our present model of the situation is (possibly) correct. The trouble is that the present model cannot be correct. For consider the position of Yahweh at t — 1, a minute (say) before he acts sinfully. Since Yahweh occupies D at $t-1$, Yahweh is at $t-1$ omnipotent, omniscient, and without sin. Being all-knowing at $t-1$, Yahweh must then realize that he is about to act sinfully and so is about to vacate his divine office. Being omnipotent at $t-1$, Yahweh has it within his power to see that things do not turn out this way. After t Yahweh can be charged with having *foreseen* that he was about to commit a sinful action—one that it was within his power not to perform. In light of this, I doubt that Yahweh can be sinless at $t-1$, and so doubt that he can occupy D at $t-1$. One surely is guilty of *some* sort of moral failing in the event that one realizes one is about to act maliciously (say), it is within one's power not to so act, and yet one proceeds to act maliciously. In such cases the moral transgression lies, not merely with what one does (the malicious action), but with what one does not do (namely, decide not to act). Accordingly, I am sceptical of the idea that there is such a world as W_s, in which an individual ceases by way of a moral failing to occupy the divine office. No possible world is such that one of its inhabitants sins at t but also is omnipotent, omniscient, and 'wholly good' at $t-1$.

In light of this fact, I believe that Pike's thesis that the requisites for the divine office are consistent should be rejected.[4]

North Carolina State University,
Raleigh

© W. R. CARTER 1982

---

[4] Comments by the Editor and Mark Richard have helped to eliminate mistakes in earlier drafts of this paper.

CAUSAL DEVIANCY AND MULTIPLE INTENTIONS

By James A. Montmarquet

SUPPOSE that S's intention to do X is causally connected to his X-ing only via the intention of another agent—only via another agent's intention that S do X. Does this imply that S cannot have *intentionally* done X? The question, we shall see, is important for the problem of trying to give a causal account of intentional action.

When a person both X-es and intended to X, we will say that his X-ing *matches* the intention he had. The problem, then, of giving a causal account of intentional action is that of specifying the right sort of causal chain connecting an intention to do X to matching behaviour for the latter to constitute an intentional doing of X. Not just any chain, we know, will do. When the chain is 'deviant'—as when Chopin's intention to play the piano as if nervous makes him sufficiently nervous that he unintentionally plays nervously—an intention to do X causes matching behaviour, yet no intentional X-ing occurs. On Christopher Peacocke's currently influential approach (see, e.g., his *Holistic Explanation*, Oxford, 1979, ch. 2, including p. 56 for the latter Chopin example), this distinction between deviant and nondeviant causal chains is made out in terms of the difference between mere causal explanation and what Peacocke calls 'differential explanation'. In the deviant case, a causal explanation will fail as a differential explanation (roughly) because the cause and the effect cannot be described in such a way that the laws connecting them (as so described) specify the one as a *function* of the other. So it is Peacocke's contention that in the normal (nondeviant) case we can describe the agent's intention and matching behaviour in such a way that they exhibit the required sort of functional dependence; in the deviant case we cannot; and his contention, further, that such functional dependence captures our intuitive idea that in the deviant case it is a rather singular accident that an intention causes matching behaviour—an accident in that somehow the behaviour lacked a certain kind of 'sensitivity' (p. 61) to the intention.

Peacocke himself, however, offers a case in which this differential explanation ('DE') requirement is met by a *deviant* causal chain, that of a

> (1) knowledgeable neurophysiologist who decides on a particular occasion to produce in me exactly the motor impulses to realize what he knows, from my neurophysiological states, to be my intentions (p. 87).

In recognition of this, Peacocke augments the DE requirement with one that 'the chain from intention to bodily movement should not run through the intentions of another person' (p. 88), thus yielding what he believes to be a sufficient as well as a necessary condition of intentional action.

John Bishop ('Peacocke on Intentional Action', ANALYSIS 41.2, March 1981, p. 96) has objected that this augmentation, formulated as it is directly in the language of our intuitions, specifically precludes Peacocke from offering what he intends: an explanatory account of what those intuitions were about. Peacocke's aim, Bishop writes, was to 'supply a theoretical account of the deviant/nondeviant distinction which will make precise, and simultaneously unify, the vague and metaphorically expressed intuitions we have about it'; but with the added 'ungainly exclusion condition expressed in the language of our immediate intuitions', says Bishop,

> the account becomes lopsided . . . Peacocke's strategy is under threat. For, in the light of this concession, why should we retain the confidence with which Peacocke began . . . that it is our intuitions about the 'sensitivity' of a causal chain which hold the secret of a clear understanding of intentional action, rather than our intuitions about 'origination', or 'agent-causation' (ibid.).

I shall now try to show, *contra* both Peacocke and Bishop, that this augmentation of the DE requirement is not needed in the first place—in fact, that it makes the resulting condition so strong that it no longer states a necessary condition of intentional action; and that the augmentation which *is* required will not have to be expressed 'in the language of our immediate intuitions', but will explicitly invoke the DE requirement itself. (Thus in the end Peacocke's approach, if not his augmented DE condition, is vindicated.)

I will begin with a point Peacocke stresses and Bishop, for his part (see p. 97), seems willing to concede: that the right causal chain connecting intention to matching behaviour cannot be equated with the "normal" or with whatever constitutes a "reliable" chain of this sort. As Peacocke observes, a chain of the right sort here may involve links (a) whose presence is itself highly fortuitous (p. 87), (b) which even when present are only sometimes reliable (p. 70), and which (c) needn't go on within the agent's body at all (p. 72). Indeed all of these elements are combined in an example cited by Peacocke. It involves a gunman who

(2) intends to fire his gun and whose cortex sends out the initiating event: unknown to the gunman, the nerve to his index finger has been severed, but the intended movement still takes place because the impulse produced by the initiating event attracts some lightning which itself generates the required impulse in the severed section of the nerve (p. 93).

Here let us assume, with Peacocke, that firing a gun in this way could qualify as an intentional action. And let us consider the case of G, a gunman suffering (also unknowingly) from the same severed nerve:

(3) G intends to fire his gun and his cortex sends out the initiating event. Despite the severed nerve, the intended movement still takes place because G's brain has been wired to H's in such a way that the occurrence of the initiating event causes H to form an intention that G should fire the gun and the impulse produced by this intention causes the required impulse in the severed section of G's nerve.

I should think that we would be inclined to assimilate (3) to the case of (2) rather than (1). While (1) and (2) are alike in that an intention intervenes between the original intention and the matching behaviour, the intervening intention in (3) functions analogously to the lightning in (2)—and not the neurophysiologist's intention in (1). That is, it intervenes as a mere ("blind") connecting link, and not as anything which would incline us to withdraw an imputation of agency to the person who first formed the relevant intention. What this suggests, and what I believe the discussion to follow will reveal, is that it is not the occurrence of the second intention *as such* that cancels an attribution of agency to the person intending first; it is, instead, something about the way in which this intervention takes place which is salient.

Now cases (1) and (3) differ in at least two notable regards: in (1), to use Bishop's term, the neurophysiologist 'preempts' the agent's opportunity to act by intervening before the agent is able to initiate any action at all; by contrast, in (3) the intervention merely completes an action that G has already begun for himself; second, in (1) the intervention itself constitutes an intentional action, while in (3) there is no action of H issuing from his intention. To see which of these factors is crucial, let us look at a case like (3), except that the intervention *is* an intentional action:

(4) As before, G's nerve is severed and his action is completed through an intention of H. But now H is a knowledgeable neurophysiologist who, upon recognizing G's intention and the cortical event initiating his action, is caused (by his overwhelming desire to see G's action completed) to form the intention which he knows will complete it.

Now, insofar as what is in question is who (intentionally) moved G's finger, I would be inclined to assimilate this case to (1)—that is, to claim that the occurrence of the intervening intention (and its train) makes G not the agent of the movement of his finger. To be sure, unlike the agent in (1), G has not been preempted from acting: G has succeeded, at any rate, in *initiating* the causal chain which eventuates in the firing of the gun. So in likening this case to (1) we are not suggesting that G has done nothing himself, but only that the intervening neurophysiologist, and not G, should be regarded as the agent of that part of the chain which followed his intervention. While G may have tried to move his finger and tried to fire his gun, since he succeeded in doing neither *himself*, they were not intentional acts of his.

Such a view of (4) has the virtue of consistency with the way we view more normal cases. If I do something which causes another person to fire a gun (i.e., to perform an intentional act of firing a gun), the fact that I may have initiated a causal chain which eventuated in the gun's being fired does not, of course, imply that I am the agent of the latter event. Ordinarily, I would not be so regarded precisely because the latter stage of this chain was an action of another agent. Our imputation of agency to the neurophysiologist in (4) and our denial of it to G are of a piece, then, with a perfectly common view of less singular kinds of cases.

Here it should be acknowledged that we do sometimes, perhaps most notably in cases where A orders B to do X, impute agency even when the intention of a second agent does intervene.[1] Thus if the Colonel ordered a prisoner executed, it would not necessarily be false or inaccurate to describe him as having executed the prisoner, even though a subordinate actually pulled the trigger. But still, one might go on to ask whether the Colonel actually executed the man himself or only *had* him executed. With such locutions, we distinguish the "doings" of executives from the often more taxing performances of the rest of us. So I believe the point to be made here is that if (as I assume) one wants an account of intentional action which is sensitive to the distinction between doing X yourself and having X done for you by another—sensitive in such a way that only the former is treated as a full-fledged intentional action—then we remain on solid ground in maintaining that the intervening neurophysiologist, and not G, performed the intentional act of moving G's finger in (4).[2]

With that qualification, then, it seems right to say that the occurrence of a second intention, *if* it issues in an intentional X-ing, will defeat the imputation of intentional X-ing to the person whose intention initiated the causal chain in question. And this, certainly, explains why there was no such defeating of G's agency in (3), where the second intention was a "freak" occurrence which did not lead to an intentional action.

Observe, though, one other feature of the relation between the second intention in (3) and G's subsequent behaviour: the causal chain connecting them is itself a deviant one—and would be so classified by its evident inability to meet the *un*augmented DE requirement. Thus what emerges from a consideration of these cases is that Peacocke's DE requirement is capable of playing a more central role in the explanation of intentional action than perhaps he himself has recognized. For the only causal chains which clearly pose a problem for the unaugmented DE standard appear to be those which, like the ones in (1) and (4), include a second intention and matching behaviour which themselves *pass* that standard. In other words, the underlying truth here may simply be that:

---

[1] I owe this point to the ANALYSIS referee.

[2] Should we want an account which makes this distinction? The short answer, I believe, is that while theories of individual (or group) *responsibility* may indeed be interested in explicating a broader sense of intentional action (inclusive of actions done for one), presumably such theories would build upon accounts of intentional action in a narrower sense. As I read them, most recent philosophical accounts of intentional action fall into this latter category.

> The right causal chains connecting an intention to do X to matching behaviour are those which (a) meet the DE requirement and (b) have no proper part consisting of an intention to do X and matching behaviour which themselves meet the DE requirement.

Against this more modest alteration of the DE requirement, Bishop's charges lose much of their force. First, the theoretical imbalance he criticizes has disappeared now that the class of exceptions to the DE requirement has itself been partially specified in terms of that requirement. No longer do we have an explanatory (DE) account ungracefully coupled to an apparently *ad hoc* restriction not formulated in the language of the explanatory account. Second, Bishop's suggestion that, for all Peacocke has shown, perhaps there lurk more basic intuitions (concerning 'origination' or 'agent-causation') which may ultimately take over the central role Peacocke accords to intuitions concerning the sensitivity of causal chains—this possibility gains no support from the revised theory. True, it concedes that not all our intuitions about intentional action pertain to the sensitivity of causal chains: there is, too, the intuition that the agent of an event is the *last* person whose matching intention is nondeviantly connected to it. But since the latter point—as the qualifying term, 'nondeviantly', reminds us—is not independent of sensitivity considerations, we should hardly expect, on the basis of *this* account of intentional action, that intuitions besides those pertaining to causal sensitivity will supplant the latter. For on this account, though not on Peacocke's original, these other intuitions play no independent role vis-à-vis sensitivity ones.

Loyola University of Chicago © JAMES MONTMARQUET 1982

IMPERCEPTIBLE PARTS

By A. J. Gottlieb

COMMON sense and physics are often described as giving rise to different conceptions of one and the same material object. Physics attempts to explain an object's possession of its phenomenal properties such as colour in terms of its imperceptible constitution, but philosophers of perception claim to find a difficulty in this explanation. If the scientist's story is held to be literally true the question arises of how a collection of imperceptible molecules, say, can be identical with the smooth brown table I see. And if the two are not identical, what is the relation between them?

A solution Ayer once defended is that the two conceptions be fused (*The Central Questions of Philosophy*, London, 1973). The proposal is

> to conceive of physical particles as being what Locke called the minute parts of perceptible objects, in which case their being imperceptible would not be a necessary part of their nature, but simply an empirical consequence of their being so minute. On this view, it has just to be accepted as an empirical fact that particles which are individually colourless compose coloured objects when enough of them come together . . . (p. 110).

He has subsequently rejected this way out ('Replies' p. 296 of *Perception and Identity*, ed. G. Macdonald, London, 1979), but it seems to me that the only identifiable one of the arguments that dissuaded him is invalid and that with a slight modification this answer is the right one. The argument in question is Strawson's:

> If the impossibility of perceiving the ultimate components is to be viewed as merely empirical, we can sensibly ask what the conceptual consequences would be of supposing that impossibility not to exist. The answer is clear. Even if there were something which we counted as perceiving the ultimate particles, this would still not, from the point of view of scientific realism, count as perceiving them as they really are. And nothing could so count; for no phenomenal properties we seemed to perceive them as having would figure in the physical explanation of the causal mechanisms of our success. ('Perception and its Objects' p. 58, in Macdonald, ed.)

The charge, I take it, is one of vicious regress. If we were to perceive the elements physics ascribes to the table in the way in which we do in fact perceive the table, as having phenomenal properties, the original problem of reconciling the scientific and perceptual views of the object of perception would recur with respect to the miraculously enlarged molecules and whatever it was that causally explained our seeing them. Maybe we can sometimes be said to see simultaneously a feature or constituent of an object and also that feature whose presence causally explains our ability to see the other. Presumably, a special problem is thought to arise when the explanatory feature is to be the means of a comprehensive account of our perception of the phenomenal properties of the object, mentioning all the relevant things and processes that can be ascribed to the object itself. It would be said that such a feature could not be seen. Then, assuming that every perception is of, though not just of, phenomenal properties, one would never be able to see the explanatory features or constituents, except in the irrelevant sense deriving from their causal contribution to what is seen in the strict sense. So, the argument would finish, we could not under any circumstances see the scientific components of things and thus the idea that such components are the 'minute parts' of perceptible things becomes mysterious.

It may be noted that although it is indeed difficult to see how all the relevant aspects of an act of perception including reactions of the eye and brain and events in the medium could figure in the content of a perception, it is not clear why the explanatory features, as I have called them, of the object itself could not all be seen at once. But the point I wish to urge is that Strawson's conclusion is derived fallaciously anyway. Even if it is true that in every perception some feature of the object is left out, it does not follow that there is some one feature which is left out of every perception. So even if under the imagined circumstances in which molecules are perceptible something else is not, we do not have a reason for supposing that some one and the same element of the scientific story must be left out of every possible perceptual story, only that no one perceptual story can tell the whole scientific story. And why should the theory that the scientific account of perception yields what are in fact the parts of everyday things be committed to the proposition that all these parts can be seen at once, given the right conditions? All that follows from Strawson's idea that the original problem would reappear in Ayer's thought-experiment is that a persistent critic could press us to perform an unending series of similar thought-experiments to flesh out the claim that the scientific story concerns the parts of perceptible objects. The old solution may be applied at each new level. Any viciousness in this regress would seem to be purely on the part of the critic.

It might be thought that a difficulty would crop up with absolutely fundamental items if there were such things, but this is not so. On the supposition that these were seen, their apparent phenomenal properties might be explained by reference to larger items, that is, their phenomenal properties might be macro-reducible; this does not seem to be ruled out in advance. In general, we need a distinction between the objects, events, or whatever, that physics takes to figure in the final explanation if such there be of the properties of a material object, and the properties which would figure in the explanation of any perception of the former's apparent phenomenal properties. This need not be a final explanation in the previous sense: no-one thinks that the account we in fact give of the phenomenal properties of tables must make use of the ultimate concepts of physics.

But a slight modification of the leading idea is called for. Since we cannot say with confidence exactly what would happen if my table were to grow to such a size that its molecules were as large as some easily seen objects, the fact that we cannot see molecules may, for all we know, be not merely a consequence of their being very small but of some other physical facts as well. It is possible that some counterfactual suppositions would have to be made apart from those concerning size: one may have to imagine the values of certain basic constants to be changed, for example. But so long as the particles may be said to remain in the same relation to the table throughout all these fantastic changes the thesis that scientific particles and so on are the imperceptible parts of perceptible objects may be maintained. The difficulties for one sensitive to physical theory in supposing them to be perceived may be great but, as Strawson implies, it is the supposed conceptual difficulties which set up the philosophical problem.

University College London

© A. J. GOTTLIEB 1982

Philosophy Documentation Center

Bowling Green State University
Bowling Green, Ohio 43403 USA

(419) 372-2419

The Philosopher's Index
An up-to-date quarterly index with abstracts of philosophy books and journals. An annual cumulative edition is also published. In addition, two retrospective indexes have been published, i.e., the *Retrospective Index to U.S. Publications From 1940* and the *Retrospective Index to non-U.S. English Language Publications From 1940.*

Dialog
A computerized search service which provides custom bibliographies with abstracts on any logical combination of philosophical topics.

Philosophy Research Archives
The microfiche journal of philosophy which publishes articles, translations, monographs, and bibliographies.

Teaching Philosophy
Explores ideas about teaching and learning philosophy by publishing articles, discussions, and reports on new courses, methods, materials and their impact on philosophy. Textbook and audio-visual reviews are regular features.

American Philosophical Quarterly
A scholarly journal which publishes articles by philosophers of any country on any aspect of philosophy. (Formerly published by Basil Blackwell, Oxford)

Bibliographies of Famous Philosophers
Hardbound bibliographies have been published on Edmund Husserl, Jean-Paul Sartre, Alfred North-Whitehead, and Henri Bergson. Forthcoming bibliographies include works on Martin Heidegger and George Santayana.

Directory of American Philosophers
Covers the United States and Canada with information on faculties of philosophy, names, addresses, and specialties of philosophers, data on colleges and universities, philosophical societies, centers, institutes, journals, publishers, and data on fellowships and assistantships. Statistics on the profession. 1982-83 edition published January 1982.

International Directory of Philosophy
Covers Europe, Central and South America, Asia, Africa, and Australia. This directory provides names, addresses, and specialties of philosophers as well as data on colleges, universities, philosophical societies, institutes, journals, and publishers of philosophical works. 1982-85 edition published January 1982.

Directory of Women in Philosophy
A handbook of women in philosophy, with information on fields of specialization, degree (date and institution), dissertation or publications, position, institution, dossier source, address, and telephone number. 1981-82 edition published February 1981.

Mailing Labels
Labels of approximately 9,600 philosophers in the United States and Canada are available as well as addresses by specialty, i.e., Ethics, Aesthetics, etc.

Journal Typesetting
Journals are economically typeset using computer driven photo-composition equipment. Books, ads, and other documents can also be done.

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

PUBLISHER. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. US Mailing Agent, Expediters of the Printed Word Ltd, 527 Madison Avenue, Suite 1217, New York, NY 10022, USA. Second Class postage paid at New York.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £11.00 (inland), £13.50 (overseas), $33.00 (USA), $39.50 (Canada); for individuals £7.00 (inland), £8.60 (overseas), $21.00 (USA), $25.00 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year—in January, March, June and October. Orders should be sent to the publisher.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Set printed and bound by Billing and Sons Limited, Guildford and London

Vol. 42 No. 3 (New Series No. 194) June 198[illegible]

ANALYSIS

Edited by
CHRISTOPHER KIRWAN

CONTENTS

ANALYSIS competition 'Problem' No. 19
More average utilitarianisms T. M. HURKA
In defence of Lewy PETER SMITH
Wittgenstein's operator *N* ROBERT J. FOGELIN
More on Wittgenstein's operator *N* P. T. GEACH
On a possible misinterpretation of Kripke's semantics for intuitionistic logic ALLEN HAZEN
Sets whose members might not exist THOMAS BALDWIN
Internal structure and essence A. C. GRAYLING
Predicative and constructive knowledge D. E. OVER
Goldman on what justifies belief DONALD E. STAHL
Manipulability theory and event types DOUGLAS EHRING
Reply to Lowe on actions JENNIFER HORNSBY
Materialism and memory ANDREW WARD
Williams on 'The Self and the Future' HAROLD W. NOONAN
Mince pie reasoning GERALD W. BARNES
Intentionality: spontaneous ascription and deep intuition KIM DAVIES
Against a deontic argument for God's existence PATRICK GRIM
On being a cat E. J. LOWE
A thesis refutable by a sentence verifiable by its use JENNIFER HORNSBY

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003–2638

Philosophical Studies

An International Journal for Philosophy in the Analytic Tradition

Editor-in-chief: JOHN POLLOCK

Associate Editor: KEITH LEHRER

Philosophical Studies, a journal devoted to work in analytic philosophy, publishes articles in all general areas of philosophical investigation, philosophy of science, epistemology, metaphysics, ethics, and so forth, including the history of philosophy. The main criterion of articles published is that they contain rigorous argumentation and precise analysis of key concepts. Articles in the journal are intelligible to philosophers whose expertise lies outside the subject matter of the article. It is intended that a diligent reader of the journal will be kept informed of the major problems and contributions of contemporary analytic philosophy.

Subscription Information: 1982, vols. 41–42 (6 issues). Private: Dfl. 45,– / approx. US $ 18.00 per volume. Institutional: Dfl. 148,– / approx. US $ 59.20 per volume. ***Private subscriptions must be sent to the publishers.***

D. Reidel Publishing Company
P.O. Box 17, 3300 AA Dordrecht, The Netherlands
190 Old Derby St., Hingham, MA 02043, U.S.A.

The Philosophical Quarterly

Vol. 32 *No.* 128 *July* 1982

Special Number on Scientific Realism

Three Kinds of Scientific Realism	*Hilary Putnam*
Scientific Realism and Scientific Change	*John Worrall*
The Miracle of Science	*James Robert Brown*
Verification, Perception and Theoretical Entities	*Michael Luntley*
CRITICAL STUDIES	
Constructive Empiricism *versus* Scientific Realism (Bas van Fraassen, *The Scientific Image*)	*Alan Musgrave*
The Present State of Realism (Richard Healey (ed.), *Reduction, Time and Reality*)	*Graham Nerlich*
BOOK REVIEWS	BOOKS RECEIVED

The Editors intend to devote one issue in **1983** to articles on **Medical Ethics** and one in **1984** to articles on **Frege.** Contributions should be submitted, **by 31 December of the preceding year,** to The Executive Editor, The Philosophical Quarterly, The University of St. Andrews, Scotland.

Annual Subscriptions:
Individuals £8.00 ($20.00) Institutions £16.00 ($40.00)
Single issues: £4.00 ($10.00)

for orders placed directly with the Journals Manager, Scottish Academic Press, Ltd., 33 Montgomery Street, EDINBURGH EH7 5JX, Scotland.

ANALYSIS 42.3 JUNE 1982

ANALYSIS "PROBLEM" NO. 19

The nineteenth problem is set by Mr Colin Radford of the University of Kent at Canterbury, and is as follows:

If I am looking at myself in a mirror I am directly facing, do I see myself looking at myself? If so, do I also see myself looking at myself looking at myself – and so on?

The word limit is 600 words. Entries should reach the Editor of ANALYSIS by 30 November 1982; they should not be sent to Mr Radford. Entries will not be returned unless accompanied by stamps or international postage coupons. Contributors may submit entries under their own names or a pseudonym. Contributors must be under the age of thirty, or undergraduates or graduate students.

A report with any winning entries will be published in volume 43 of ANALYSIS. It has been agreed on behalf of the ANALYSIS Committee that £75 will be available to be awarded as a prize if Mr Radford finds a sufficiently deserving contribution.

MORE AVERAGE UTILITARIANISMS

By T. M. HURKA

IN 'Average Utilitarianisms' (ANALYSIS 42.2, March 1982, p. 65) I distinguished six different average utilitarian theories and showed that, although they are all open to powerful objections, the objections are in many cases importantly different. These six theories agree in calculating value, which is what they tell us to maximize, by calculating happiness across all the persons and times in human history. But they disagree both in the principles they use to make this calculation, and in the order in which they apply them. Two theories (which I called A1 and A2) use averaging principles to calculate happiness across persons and summative principles to calculate happiness across times, while another two (A3 and A4) use averaging principles both across persons and across times, and another two (A5 and A6) use summative principles across persons and averaging principles across times. Three theories (A1, A3 and A5) apply their principles across times before their principles across persons, while another three (A2, A4 and A6) apply their principles across persons before their principles across times. It was one of the main points of 'Average Utilitarianisms' that, once we start using averaging (or similar) principles to calculate happiness, the order in which we apply these principles takes on a significance which it does not have in classical utilitarianism.

In this paper I want to distinguish five more average utilitarian theories. These theories calculate happiness using the same principles in the same order as A1, A3, A4, A5 and A6, but they apply these principles across only persons and times in the *future* of human history. Now in classical utilitarianism, and even in a theory like A2, which finishes its calculations of happiness by summing across times, the difference between calculating happiness across all persons and times and calculating it across only future persons and times is of no practical significance. Since what we do in the present cannot affect the amount of happiness felt in the past, the only way that we can increase the total happiness in human history is by increasing the total happiness in the future, and this means that the practical import of the two calculations will always be the same. But the difference becomes significant once we start using averaging (or similar) principles across times or lives.[1] In 'Average Utilitarianisms' we saw that A1 and A3–A6 are not 'forward-looking' in the way classical utilitarianism is, but sometimes allow the rightness and wrongness of actions to be affected by facts about the past. A1, A3, A4 and A6, for instance, sometimes require us to let the human race die out just because people will be a little less happy in the future than they were on average in the past. And A3 and A5 sometimes require us to kill a person just because he will be a little less happy in the future than *he* was on average in the past. Average utilitarian theories which calculate happiness across only future persons and times must by their nature be forward-looking, and this may encourage the thought that these theories, which we can number A7–A11, will be more attractive than A1–A6.

Although they are certainly forward-looking A7–A11 are open to many of the same objections as A1–A6. To see this, let us examine what some may consider the most plausible of these theories, namely A7. Since A7 uses the same principles in the same order as A1, it tells us to maximize the average of the total happiness in every life (or part of a life) that will be lived from now on into the future. A7 has in the first place some unattractive consequences about letting the human race die out which are similar to those of A1. Although A7 does not tell us to let the human race die out if people will be less happy in the future than they were in the past, it does tell us to let the human race die out if people will be less happy in the distant future than they will in the near future. Imagine that everyone alive in the world today has an extremely happy life ahead of him, and will enjoy a total of 100,000 units of happiness between now and his death (this on a scale where the best we do today is 1,000 units of happiness in a life). But any people born after today will enjoy only 90,000 units of happiness in their lives. Then even though these future people will lead very happy lives A7 says we must not produce them, but let the human race die out. A7 also has some unattractive consequences about killing. Imagine that almost everyone alive in the world today will enjoy 100,000 units of happiness between now and his death, and that future people will also enjoy 100,000 units of happiness in their lives. But one person alive today will enjoy only 90,000 units of happiness between now and his death. Then A7 tells us to kill this person. His survival would lower the average happiness per life (or part of a life) lived from now into the future, and we must therefore prevent it. The killing of below-average people is also a requirement of A2, but it is important to realize that A2 and A7 interpret 'below-average' somewhat differently. A2 requires us to kill a person if, at all or most of the times when he is alive in the future, he will feel less happiness *at those times* than other persons. But A7 requires us to kill a person if he will feel less happiness *in the remainder of his life as a whole* than other persons. Now a person may feel less than the average happiness in the remainder of his life as a whole because, while living as long as other persons, he feels less happiness than they at all or most times when he is alive. But he may also feel less than the average happiness in the remainder of his life because, while feeling as much or even more happiness than they at all or most times when he is alive, he does not live quite as long as they. Imagine that the persons who are going to enjoy 100,000 units of happiness between now and their deaths will live for 100 years and enjoy 1,000 units of happiness every year. But the person who will enjoy only 90,000 units of happiness between now and his death will live for 45 years and enjoy 2,000 units of happiness every year. Then even though at every time in the future this person will be twice as happy as anyone else, A7 (but not A2) requires us to kill him. A7's consequences about killing seem, if this is possible, even more unattractive than those of A2. Because the total happiness remaining in a person's life diminishes steadily as he gets older, A7 is committed to a systematic policy of killing the elderly, no matter how contented they may be.

---

[1] The importance of this difference for average utilitarian theories is pointed out, with an acknowledgement to some unpublished work of Derek Parfit, in Jefferson McMahan, 'Problems of Population Theory: A Review of R. I. Sikora and Brian Barry, eds., *Obligations to Future Generations*', *Ethics* 92 (1981), 96-127, p. 113. McMahan goes on to raise some interesting questions about the theory I have called A1.

There is a modified version of A7 (let us call it A7′) which avoids at least these last unattractive consequences about killing, and which has been defended by Michael Neumann in 'Killing and Average Utility', ANALYSIS 40.1, January 1980, pp. 35-6. A7′ tells us to maximize the average of the total happiness in every life (or part of a life) lived from now on into the future by persons who satisfy one of two conditions: (i) they exist now, when we are performing our actions; (ii) they will exist in the future, after we have performed our actions. Since a person we kill will always satisfy the first condition – he will always exist now, when we are performing our actions – his life will always be one of those across which A7′ has to average; and since his life will always contain 0 units of happiness from now on into the future, his death will not be required by A7′ in any circumstances where it would not be required by classical utilitarianism as well.

Although A7′ does not require the killing of below-average people it is still not a very attractive theory. The inclusion of condition (i) marks a partial departure from the forward-lookingness which was A7's main source of attraction *vis-à-vis* those other average utilitarian theories, most notably A1, which also do not require the killing of below-average people. And A7′ still has the other unattractive consequences of A7. These include its consequences about letting the human race die out, as well as some others which are shared by all of A7–A11, and none of A1–A6. It is an important feature of classical utilitarianism that it is what Derek Parfit has called a fully *neutralist* moral theory. It is both agent-neutral, in that it assigns the same moral goal to all agents, and time-neutral, in that it assigns the same goal to these agents at all times. A1–A6, which calculate happiness across all the persons and times in human history, are also fully neutralist, but A7–A11 are not. While A7–A11 are certainly agent-neutral they are not time-neutral, but assign agents a different moral goal at every different time in human history. That this departure from time-neutrality is non-trivial can be seen from the fact that it gives rise to cross-temporal moral conflicts, in which A7–A11 require persons at one time to take steps to prevent themselves from performing actions at a later time which the same theories will at that later time be telling them they are required to perform. Let us once again take A7 as our representative of these theories. Imagine that everyone alive in the world at one time (call it *t*) will enjoy 100,000 units of happiness between *t* and his death, but that anyone born after *t* will enjoy only 90,000 units of happiness in his life. Then if it is applied at *t* A7 will say that the best outcome is one in which none of these people ever have any children, and will tell them to take whatever steps are necessary at *t* to prevent themselves from having children either at *t* or at any time subsequent to *t*. But now imagine that a certain amount of time has passed, so that, although exactly the same people are alive in the world, they have only 10,000 units of happiness to enjoy before their deaths (any children they produce will still enjoy 90,000 units of happiness in their lives). If it is applied at this later time (call it *t*′) A7 will say the best outcome is one in which these people all have as many children as possible, and will tell them to perform one of the very actions, namely having children at t', which at the earlier time t it was telling them to do everything in their power to prevent.

That A1 and A3–A6 are time-neutral but not forward-looking, while A7–A11 are forward-looking but not time-neutral, is a matter of some importance. Because classical utilitarianism is both forward-looking and time-neutral, many philosophers have assumed that these two formal features are characteristic of consequentialist theories in general. We have seen, however, that in theories which use averaging (or similar) principles across times or lives the two features are actually incompatible. If these theories calculate happiness across all the persons and times in human history they can be time-neutral but they cannot be forward-looking; whereas if they calculate happiness across only future persons and times they can be forward-looking but they cannot be time-neutral.

University of Calgary, Calgary
Alberta, Canada T2N IN4

©T. M. Hurka 1982

IN DEFENCE OF LEWY

By Peter Smith

IN his book *Meaning and Modality*, Casimir Lewy argues that our preformal use of the concept of entailment is governed by a number of intuitive constraints which cannot in fact all be satisfied together. He concludes that the various 'analyses' of entailment to be found in the literature are best regarded as alternative reconstructions of an inconsistent concept.

Our assessment of such a reconstruction will obviously depend on the extent to which it violates our preformal intuitions. And some proposed reconstructions can be seriously faulted for being unnecessarily revisionary and requiring us to abandon too many of our initial intuitions. Lewy criticizes the Anderson–Belnap theory of entailment in just this way. He agrees, of course, that their entailment calculus formalizes an implication relation which avoids the so-called 'paradoxes of strict implication'. But, like many others, Lewy thinks that the Anderson–Belnap rejection of disjunctive syllogism is far too high a price to pay for avoiding those paradoxes.

Lewy finds further reasons for dissatisfaction with the Anderson–Belnap approach. In discussing four paradoxes of his own devising, he claims that their theory cannot be extended in any natural way to deal with these new paradoxes without again violating deeply embedded intuitions. So, quite apart from the difficulty over disjunctive syllogism, the theory still does not provide the basis for a general paradox-free account of entailment. However, this claim has recently been criticized by John K. Slaney in his paper '"Entailment" survives Lewy's paradoxes' (ANALYSIS 41.4, October 1981, pp. 181–91). He argues that Lewy's *second* line of attack on the Anderson-Belnap approach 'is easily repulsed'. But not so.

Consider the following two theses about entailment which both have a very considerable intuitive appeal. The first thesis is not challenged by Slaney, so I will just repeat it here without further comment:

(I) If A entails B and B entails C, then A entails C.

The second thesis can be introduced by means of a couple of examples. Take the propositions

P John is a father
Q All fathers are parents
R John is a parent.

Then undoubtedly ⌜P & Q⌝ entails R; indeed this entailment holds in virtue of the principles of first-order logic. But it is also surely the case that P taken by itself entails R. Of course, this latter entailment is not ratified by elementary logic alone; but then it is not to be supposed that entailments are restricted to those that can be so ratified. On the contrary, the claim that P entails R is itself the very paradigm of an intuitively acceptable entailment claim. Again, take the propositions

P′ There are three mothers and four fathers in the room
R′ There are seven parents in the room.

Surely P′ entails R′. Of course, to get an entailment that is ratified by elementary logic, we would have to conjoin with P′ the propositions that all mothers are parents, all fathers are parents, no mother is a father, and perhaps (depending on how we regiment the numerical quantifiers) the proposition that three plus four is seven. Call the conjunction of these necessary propositions Q′. Then evidently ⌜P′ & Q′⌝ entails R′. But, to repeat, it seems to be equally acceptable to say simply that P′ entails R′.

In both our examples, the underlying principle seems to be this: when we are arguing about purely contingent matters (e.g. about whether John is a parent, or the number of parents in a given room), then we can legitimately suppress necessarily true premises. In other words, failure to mention necessarily true premises whose addition would give us an entailment ratified by elementary logic does not falsify an entailment claim. Suppose we call a proposition *thoroughly* contingent if it is contingent and is not a truth-function with some non-contingent constituent. Then the principle exemplified by our examples seems to be captured by the following thesis:

(II) If ⌜A & B⌝ entails C, where A and C are thoroughly contingent and B is necessary, then A entails C.

On reflection, this thesis seems intuitively acceptable.

Now, Lewy's first three paradoxes show that in fact the plausible theses (I) and (II) quickly lead to paradox. So a paradox-free general theory of entailment must either restrict transitivity or circumscribe more tightly than (II) the cases where a necessary premise can be dropped. But transitivity is built into the foundations of the Anderson–Belnap approach: so if an extension of their theory is to deal with Lewy's paradoxes it must emend (II). Lewy comments:

> I do not see how they could in *general* distinguish, in a language as rich as that of my Paradox I, between the kind of logically necessary premiss that *can* be legitimately suppressed and the kind that cannot. (*MM* 131, fn. 2.)

Slaney, however, maintains that Lewy's challenge here can be easily met: the theorist in the tradition of Anderson and Belnap will just say that it is *never* permissible to drop a premise on the grounds that it is necessarily true. (II) and its variants are unacceptable and Lewy's paradoxes just do not arise.

But to this Lewy would reply, surely correctly, that the proposed ban on suppressing necessary premises is as counter-intuitive as the rejection of disjunctive syllogism. It seemed absolutely clear in our examples above that P entails R, and P′ entails R′. And we are inclined to say that, should one be pressed for a justification of these entailment claims, then one could refer to the facts that ⌜P & Q⌝ entails R, and ⌜P′ & Q′⌝ entails R′, where Q and Q′ are necessary. If this sort of justification is not accepted one must either say (a) P does *not* entail R, and P′ does *not* entail R′, or (b) P entails R, P′ entails R′, but these and the myriad other entailment claims like them are all primitive entailments which cannot be further justified. Neither alternative seems very plausible.

Now, perhaps the blanket ban on the suppression of necessary premises can be rescued from the charge of leading to absurdity. But certainly, it is not an adequate response to Lewy's challenge simply to propose such a ban without *any* attempt to explain away its apparently counter-intuitive consequences. For, of course, Lewy's question was not 'can an entailment theorist avoid my paradoxes?' but rather 'can a theorist avoid my paradoxes by appeal to plausible principles which do not lead to further absurdities?' Slaney's proposed ban just doesn't answer the question.

Slaney is led astray here because he has noticed an infelicity in Lewy's presentation of his own case. In discussing his paradoxes, Lewy appeals not to the thesis (II), but to (II′), which is what results from deleting the word 'thoroughly' from (II). So it is (II′) which Lewy says 'seems intuitively acceptable'. Slaney is able to show that (II′) combined with (I) leads immediately to the counter-intuitive result that a contingent proposition entails any necessary one. And Slaney takes this as showing that (II′) is after all *not* intuitively acceptable. Let us accept this. It certainly does not follow that Lewy was *completely* in error, and that nothing stands in the way of a blanket ban on suppressing necessary premises. On the contrary, as we have seen (II) *is* highly plausible, it is as strong as Lewy needs for his arguments, and it cannot be undermined by the sort of argument which Slaney deploys against (II′). Lewy perhaps misrepresents the grounds of his intuitive claims; but Slaney is wrong to jump to the conclusion that Lewy has no grounds at all.

In short, then, Lewy's first three paradoxes continue to present a problem for the Anderson–Belnap approach.

In his fourth paradox, Lewy produces a simple argument which apparently establishes that

(A) There are exactly ten brothers

entails

(B) There are exactly as many brothers as brothers

Lewy suggests that this is 'at least a *mild* paradox' (*MM* 124). But if we are not prepared to deny the transitivity of single-premise entailment, i.e. if we are not prepared to reject thesis (I) above, then it seems that we will have to learn to live with the mild surprisingness of this result. Slaney, however, disagrees: on behalf of the Anderson–Belnap theorist he suggests that (B) is at least three ways ambiguous. And he claims that on one reading (A) does entail (B), on a second (A) does not entail (B), while on a third it is the conjunction of (A) with the proposition that ten is identical to ten which entails (B). Slaney calls this 'a delicate, penetrating and intuitively satisfying analysis of paradox 4'.

I cannot imagine that Lewy would agree. Early in his book, he remarks that 'by manufacturing a non-existent ambiguity one could *pretend* to destroy *any* argument' (*MM* 9). That remark seems applicable here. For it is extremely counter-intuitive to suppose that ordinary English statements of the form 'There are exactly as many F's as G's' are equivocal. In particular, the supposition that such statements are ambiguous initially looks a lot more counter-intuitive than Lewy's result that (A) entails (B). So Lewy could quite justly claim that the attempt to side-step his fourth paradox by treating its conclusion as ambiguous again only avoids one paradox at the price of embracing a greater one. Hardly the 'triumph for Anderson and Belnap' that Slaney announces.

This point by itself might well be thought sufficient to undermine Slaney's strategy in his discussion of Lewy's fourth paradox. But there is perhaps some interest in examining the various readings which Slaney suggests for (B). Consider in particular

(B1) For some n there are exactly n brothers and exactly n brothers.
(B2) For every n there are n brothers if and only if there are n brothers.

Slaney claims firstly that (B1) and (B2) do indeed both express propositions that are also expressed by (B); and secondly he holds that (B1) and (B2) are 'classically ... simply equivalent', but do not mutually entail each other on the Anderson-Belnap theory. These claims, each of which Slaney needs for his analysis of the paradox, are both resistible.

It is not entirely clear what the intended logical form of (B1) and (B2) is supposed to be. However, both Slaney's reference to them as being classically equivalent and also other indications very strongly suggest that we are to understand (B1) and (B2) as involving classical objectual quantifications. But if so, then the ontological commitments implied by their quantifications over the domain of numbers are arguably distinct from the commitments of the original (B). For (B) is of the form 'There are exactly as many F's as G's; such propositions invite analysis in Frege's style as involving a second-order quantification, and requiring the existence of a suitable one-one relation between the F's and the G's, rather than the existence of numbers *qua* entities fit to be quantified over. If the ontological loadings of (B1) and (B2) are indeed distinct from those of (B), then neither can be accepted as representing legitimate readings of (B) after all.

Suppose we waive this first problem and move on to ask about the logical form of the expression 'there are ξ brothers' which occurs in both (B1) and (B2). Given Slaney's willingness to quantify into the position occupied by the place-holder here, we had better read this as having the form 'the number of brothers $= \xi$'. But on this understanding of (B1) and (B2), they are not even *classically* equivalent, if that is to say that they mutually entail each other by the principles of classical first-order logic with a classical theory of descriptions. For (B2) but not (B1) is classically consistent with the extra-logical proposition that there is no number which is the number of brothers. So if we accepted that both (B1) and (B2) *are* expressed by (B), then even from a classical perspective we would discern an ambiguity. Further, we could note that while (B2) is a theorem of pure quantification theory with descriptions, establishing (B1) requires the extra-logical premise that there is indeed some number which is the number of brothers, which can be shown by appeal to (A). So even from a classical perspective we could make good sense of the idea that (A) is relevant to (B1) in a way that it is not relevant to (B2): (A) is not needed to establish (B2), while it – or some substitute – *is* needed to classically establish (B1).

In short, if we *did* accept the contentious claim that both (B1) and (B2) formally express propositions expressed by (B), we would not need to appeal to the Anderson-Belnap theory of entailment in order to give a Slaney-style diagnosis of Lewy's fourth paradox. On the contrary, from a purely classical perspective we could still say that (B) was ambiguous between two propositions, one of which followed from (A) and *needed* some such extra-logical support, while the other was naturally proved by pure logic without appeal to (A). This seems just as 'penetrating and intuitively satisfying' an account of our wavering intuitions about the paradox as the account which Slaney offers, so there is certainly no special 'triumph' for the Anderson-Belnap theory here, even if we overlook the implausibility of deeming (B) to be ambiguous in the way Slaney suggests.

In summary, to return to the central point: Slaney is correct in claiming that the Anderson-Belnap approach *can* defuse Lewy's four paradoxes by blocking the arguments; but this claim by itself does not meet Lewy's challenges. For Lewy can still maintain that, just as Anderson and Belnap avoid the Lewis paradoxes at the absurd price of rejecting disjunctive syllogism, so they can only avoid his own paradoxes by yet further stretching our credulity to breaking point.

University College of Wales,
Aberystwyth, Dyfed SY23 3DY

© PETER SMITH 1982

WITTGENSTEIN'S OPERATOR *N*

By ROBERT J. FOGELIN

'The introduction of any new device into the symbolism of logic is necessarily a momentous event.' 5.452.[1]

IN a note in this journal ('Wittgenstein's Operator *N*', ANALYSIS 41.4, October 1981, pp. 168-71) Peter Geach claims that my treatment of the Tractarian operator *N* contains a fundamental error and, for this reason, he holds that my criticisms of the logic of the *Tractatus* are unsound. Geach's challenge is ingenious. Showing that it fails will prove instructive.

Economy of exposition will be served by allowing Geach to speak for himself:

> The *N* operator yields joint denial of an arbitrary number of propositional arguments, from 1 upwards. Applied to one argument it yields the same result as ordinary negation; applied to two arguments it yields binary joint denial; but it may be applied to more than two arguments, and then these

---

[1] All references to Wittgenstein's *Tractatus Logico-Philosophicus* are to the Pears-McGuinness translation, London, Routledge and Kegan Paul, 1961.

> need not be listed, but may be given by specifying the class containing just such-and-such propositions. Wittgenstein spells all this out quite unequivocally (5.501) . . .

Up to this point there is no difference between my explanations of the operator *N* and his. This is the core doctrine as it is actually stated in the text.

Geach then, as he acknowledges, goes beyond the stated text to elucidate the underlying structure of the operator *N* in these words:

> To bring out in full the way Wittgenstein's *N* operator works, we need (something he does not himself provide) an explicit notation for a class of propositions in which one constituent varies. I shall write '$N(\ddot{x}:fx)$' to mean joint denial of the class of propositions got by substituting actual names for the variable in the propositional function (represented by) 'fx'. Thus '$(\exists x)fx$' and '$(x)fx$' will come out in a Wittgenstein-style notation as '$N(N(\ddot{x}:fx))$' and '$N(\ddot{x}:N(fx))$' respectively. The first is the denial of the joint denial of a class of propositions saying that such-and-such is *f*, and thus says that something or other is *f*: the second is the joint denial of the class of propositions saying that such-and-such is *not f*, and this amounts to saying that nothing is not *f*, i.e. that everything is *f*.

Geach concludes by showing how the *N* operator, so understood, can be used to represent the mixed multiply general proposition '$(\exists x)(x)(Fxy)$'. The construction takes the following form:

$$N(N(\ddot{x}:(N(\ddot{y}:N(fxy)))))$$

I claim that there is no way of constructing universal propositions of the form $(x)fx$ unless the system of Tractarian logic is enriched with further truth functional operators such as logical product. I further argue that even after such enrichment there will be no way of constructing such mixed multiply general propositions as $(\exists x)(x)(fxy)$. It is for this second reason, rather than the first, that I hold that the logic of the *Tractatus* is fundamentally flawed (R. J. Fogelin, *Wittgenstein*, Routledge and Kegan Paul 1976, chapter IV, §1). It is clear then that a sharp issue has been drawn between Geach and myself. If Geach's construction of these two sorts of propositions is legitimate within the framework of the Tractarian system, then my position is simply wrong.

We can begin by examining Geach's treatment of universal propositions. Geach renders '$(x)fx$' in his Wittgenstein-style notation as '$\mathrm{N}(\ddot{\mathrm{x}}:\mathrm{N}(\mathrm{fx}))$'. The surface grammar of this expression may suggest only two successive applications of the operator *N* to an initial set of propositions, but that, of course, is false. For Geach the expression '$(\ddot{\mathrm{x}}:\mathrm{N}(\mathrm{fx}))$' specifies a set of propositions that is the result of possibly infinitely many applications of the operator *N* to a set of propositions. We can contrast a construction of this kind with another in Wittgenstein's own notation: '$\mathrm{N}(\mathrm{N}(\overline{\mathrm{f}(\mathrm{x})}))$'. Here the expression '$\overline{\mathrm{f}(\mathrm{x})}$' represents the set of possibly infinitely many propositions that we get through substituting names for the variable 'x'. This set of propositions is then subjected to two successive applications of the operator *N*. Schematically, the difference can be represented in the following way:

$$\text{Wittgenstein: } N(N(\overline{f(x)})) = N(N(fa, fb, fc, \ldots)) = (\exists x)(fx)$$
$$\text{Geach: } N(\ddot{x} : N(\grave{f}x)) = N(N(fa), N(fb), N(fc), \ldots) = (x)\, fx$$

Notice that in each case we may be dealing with infinitely many propositions of the form *fx*. These propositions need not be given through enumeration and, however they are specified, they may be simultaneously brought under a single application of the operator *N*. Contrary to Geach's suggestions, I have raised no objections on this score. Our interpretations of the symbolism found in the text of the *Tractatus* are identical.

There is, however, an important difference between the constructions that Wittgenstein's notation will sanction and those permitted by Geach's Wittgenstein-style notation. In particular, and this is what gives his symbolism its force, the Geach notation allows the representation of infinitely many (non-reducible and unordered) applications of the operator *N*. Here I shall not object that this will involve the *enumeration* of possibly infinitely many applications of the operator *N*. Enumeration does not concern me. For that matter, I have no personal reservations about forming sets (and propositions out of these sets) in just the way that Geach proposes. The point is that such a procedure cannot be used to elucidate and defend the Tractarian system since it flatly contradicts one of its fundamental theses:

> 5.32 All truth functions are results of successive applications to elementary propositions of a finite number of truth operations.

Geach's notation violates the demands for the *finiteness* and *successiveness* of application found in this proposition. (The demand that the applications be successive is particularly important since it in no way suggests that we are dealing with contingent incapacities of finite creatures to carry out infinite tasks.)

My criticism of Geach's proposals can be converted into a criticism of the *Tractatus* itself: if the logical system of the *Tractatus* can be made to work only through the addition of procedures of the kind that Geach suggests, then, since it explicitly proscribes such procedures, the logical system of the *Tractatus* is fundamentally flawed.

I believe that these considerations completely dispose of Geach's attempts to employ a notational innovation to answer the criticisms that I have made against the logic of the *Tractatus*. I shall conclude with a few brief comments about the significance of this failure.

At 6.126 Wittgenstein tells us that 'one can calculate whether a proposition belongs to logic, by calculating the logical properties of the *symbol*.' This remark, together with many others (including the sequence 5.2 through 5.4), plainly indicates that Wittgenstein is committed to a decision procedure for propositions of logic.

This is how Wittgenstein has generally been read and, indeed, this seems to be one of the fundamental features of his position.[2] Since one can decide whether a proposition belongs to logic by calculating the logical properties of the symbol, the truth of a proposition of logic will not depend upon the disposition of independent logical objects. It is in this way that Wittgenstein's commitment to a decision procedure goes hand in hand with his 'fundamental idea' that 'the "logical constants" are not representatives; ... ' (4.0312).

Although Wittgenstein could not have been expected to anticipate this result, it has been shown by Alonzo Church that a decision procedure is not possible for all the propositions of first order quantification theory. In particular, no decision procedure exists for certain mixed multiply general propositions. Now Geach has produced a notation for constructing propositions which, given Church's Theorem, will not carry a decision procedure with it. So Geach must hold at least one of the following theses:

I. Despite apparent texts to the contrary, Wittgenstein did not hold that the propositions of logic admit of a decision procedure.
II. Church's Theorem is false.

His essay gives no indication which of these theses he accepts. This, needless to say, is a fundamental issue that demands clarification.

Dartmouth College, Hanover,
New Hampshire 03755, U.S.A.

© Robert J. Fogelin 1982

---

[2] Here I am in complete agreement with Max Black's claim that '... the need for known decision procedures for checking on putative logical truths is an integral and indispensable feature of Wittgenstein's philosophy of logic.' Max Black, *A Companion to Wittgenstein 'Tractatus'*, Cambridge University Press, 1964, p. 319.

MORE ON WITTGENSTEIN'S OPERATOR *N*

By P. T. Geach

THE aim of Wittgenstein's using the operator *N* was to give us a perspicuous symbolism in which the truth-functional operation was set clearly apart from a specification of the class of operands (cf. 5.521). He did not, however, show clearly in his text how the class of operands is itself to be displayed when a given propositional constituent is being picked out to be varied. In my note I showed how this can be done: the requirement that it be done in some such way is a matter of the right logical multiplicity – cf. 4.0411 on inadequate notations for generality. The notation I provided has the same expressive power as familiar quantifier-notations; so at least I claimed, and Fogelin has done nothing to upset that claim. The *Tractatus* text he has chosen as a motto may perhaps suggest that my proposed notation is a 'momentous' departure from the *Tractatus*; but that would be wrong. By Wittgenstein's way of thinking about the operator *N*, the *way* its class of operands is specified is unimportant (5.501): assuming of course that the specification is quite definite, as by my method it is.

Fogelin still appears, as in his book, to confuse the performance of *one* operation on a (possibly) infinite class of operands with the performance of an infinite number of operations.

The claim I made for my Tractarian notation was only for its *expressive* completeness. (There will incidentally be completeness of a *proof procedure* for the logical truths expressible, just as for ordinary first-order logic.) I said nothing about decidability or about Church's theorem, and Fogelin is creating confusion by mentioning these topics. His claim was to have produced quantificational formulas *inexpressible* by Tractarian devices; this claim has been refuted.

One point may have worried some of my readers: whereas the result of the operator *N* will always, as I pointed out, be a single proposition, its operand may be a single proposition or a class of propositions. But even for complicated set theories there are relative consistency proofs showing that a theory's consistency is not blemished if it is modified so as to accommodate what Quine calls *individuals*: objects that are their own one-member classes. Contradiction is certainly not to be feared if in the very rudimentary set theory of the *Tractatus* we treat each proposition as a one-member class of propositions, so that the operator *N* always operates on a class of propositions.

University of Leeds, and
3 Richmond Road, Cambridge

© P. T. GEACH 1982

ON A POSSIBLE MISINTERPRETATION OF KRIPKE'S SEMANTICS FOR INTUITIONISTIC LOGIC

By ALLEN HAZEN

THERE is no doubt of the utility of Kripke's model theoretic analysis of intuitionistic logic (cf. S. A. Kripke, 'Semantical Analysis of Intuitionistic Logic I', in J. N. Crossley and M. A. E. Dummett, eds, *Formal Systems and Recursive Functions*, Amsterdam: North-Holland 'Studies in Logic', 1965) for the classically trained logician who wishes to gain a 'feel' for what is and is not provable intuitionistically, and it is certainly not obviously foolish to hope that Kripke's account or something like it can lead to a philosophically plausible semantics that is faithful to the intentions and concerns of Brouwer and his school. In this note I will describe certain technical results about Kripke's analysis of (the propositional fragment of) intuitionistic logic that might be thought to exclude some attractive intuitive interpretations of the model theory, and will argue that they should not be taken as excluding those interpretations.

Kripke's model theory for intuitionism is the result of hybridizing the well-known interpretations of intuitionistic logic in the modal logic S4 with his own 'possible worlds' model theory for modal logic. It is itself a 'possible worlds' analysis involving a reflexive and transitive 'accessibility' relation; atomic formulas are assumed to be true at a 'world' only if true at all worlds accessible from it, the truth of conjunctions and disjunctions at a world is treated classically, but a negation (conditional) is taken to be true at a world just in case the negated formula is not true (the consequent is true if the antecedent is) at any accessible world. The problem of finding an intuitive interpretation, then, is that of interpreting the notions of a world and of accessibility.

Kripke's own proposal is that the worlds be thought of as representing 'points in time (or "evidential situation"), at which we may have various pieces of information...' The truth of a formula at a world is taken to represent our having enough information, in that 'evidential situation', to prove a proposition: the formulas true at a world represent the propositions that have been verified by that point in time. One world's being accessible from another, finally, is taken to correspond to its representing a possible later 'evidential situation': perhaps (since accessibility is reflexive) a situation in which we have no more information, but perhaps one in which we have been able to prove additional propositions.

Grzegorczyk (cf. his 'A Philosophically Plausible Formal Interpretation of Intuitionistic Logic', *Indagationes Mathematicae* 26 (1964), pp. 596-601) arrived independently at a similar interpretation of a formal semantics which is, as far as propositional logic is concerned, equivalent to Kripke's.

Reflection on the semantic paradoxes, however, and in particular reflection on Poincaré's notion of predicativity, and his suggestion that that reasoning is predicative which is not affected by an enlargement of the universe of objects considered, leads to a different, less blatantly epistemological, interpretation. We may think of the 'possible worlds' as different universes of mathematical entities, one 'world' being accessible from another if it is an expansion of it. We may think of some basic mathematical objects, such as the natural numbers, as belonging to all of the worlds, but the domains of 'higher type' objects, such as properties and functions, will in general vary. The move from one world to an accessible one, then, corresponds to the discovery or invention of new methods of defining functions and properties. The move to another world can yield new information, in the sense of making possible the proof of new theorems about objects already defined, through the application of such general principles as mathematical induction to the newly defined properties. Alternatively, if the formulas interpreted in a model are thought of as representing assertions about mathematical entities in general rather than, e.g., number-theoretic statements, formulas may be thought of as 'coming true' in the move from one world to an accessible one as the examples needed to make them true come into existence. Thus, for example, the statement that there are functions with such-and-such properties might not be verified at worlds representing fairly elementary means of definition but become true when the universe of functions is expanded enough to include some with the desired properties.

A problem arises, or seems to arise, for both of these intuitive interpretations because it is at least plausible to suppose that their accessibility relations have a formal property which, according to a well-known model-theoretic result, should prevent them from being models for intuitionistic logic.

The formal property is *directedness*: for any pair of possible worlds, there is some possible world which is accessible from both. It is at least plausible to suppose that 'possible evidential situations' are a directed class under the appropriate relation. Consider two statements, each of which can be established as a theorem. Then their conjunction can also be established as a theorem: proving one of them does not invalidate the proof of the other, so after establishing one it will still be possible to establish the other. This, however, comes close to the claim that possible evidential situations are a directed class: obtaining the new mathematical information of one situation will not make the new information of the other unavailable, so it is possible to progress to an evidential situation in which we have all the information of the two combined. To avoid this conclusion one must suppose that *incompatible* advances in our mathematical knowledge are possible: that it may, for example, be possible both to establish *A* as a theorem and to establish *Not-A* as a theorem, for some proposition *A*. It is unclear what sort of possibility is at issue here, but on at least some obvious construals, both classical and intuitionistic, it is absurd to suppose that both the establishment of *A* and the establishment of *Not-A* are possible extensions of our mathematical *knowledge*. (Kripke tries to avoid this by suggesting that the possibility is epistemic: 'If *H* is any situation, we say that [*H'* is accessible from *H*] if, as far as we know, at the time *H*, we may later get enough information to advance to *H'*'. It may be possible to make this notion of epistemic possibility precise, but, at least in the cited article, Kripke does not further analyse the notion of possibility 'for all we know'.)

The directedness of the accessibility relation on the other intuitive interpretation is if anything more evident. If two ways of

Philosophical Investigations

Edited by **D. Z. Phillips**

Philosophical Investigations is a quarterly journal, publishing articles which follow the approaches pioneered by J. L. Austin, Gilbert Ryle and Ludwig Wittgenstein, or which explain and clarify their philosophical thinking.

Quarterly from January 1982

Since its first publication in 1978, *Philosophical Investigations* has been edited in the United States by Frank Mosedale. From January 1982 it is published by Basil Blackwell, under the editorship of D. Z. Phillips, Professor of Philosophy at University College, Swansea, and with a new editorial board.

Articles in the first issues include:
Norman Malcolm, *The Relation of Language to Instinctive Behaviour*
Cora Diamond, *Anything but Argument*
Peter Winch, *Text and Context*
Jenny Teichman, *Pacifism*
H. A. Nielsen, *How Language Exists: A Question to Chomsky's Theory*
Frank B. Ebersole, *Stalking the Rigid Designator*
Raimond Gaita, *Better One than Ten*
O. K. Bouwsma, *Makes it True*
H. O. Mounce, Donald Evans, John McDowell, *Reason and Action*

Don't miss subscription rates and order form overleaf

Basil Blackwell Publisher · Oxford · England

Philosophical Investigations

Order Form

Subscriptions:
Institutions: £19.50 (UK); £23.75 (Overseas); $44.00 (US);
$52.50 (Canada)
Individuals: £10.00 (UK); £12.25 (Overseas); $22.50 (US);
$27.00 (Canada)
Rates include inland or accelerated surface postage

(PLEASE PRINT)
Name ....
Address ....
....
....
Signature .... Date ....

Please enter my subscription for *Philosophical Investigations*

I enclose cheque/international money order £ ....

Please charge my Access/American Express/Barclaycard/
Diners Club/Mastercharge/Visa account number ....

Please charge my account, number ....

Please send this form with your remittance (if applicable) to your usual subscription agency or bookseller, or direct to:

Sue Dommett, Journals Department,
Basil Blackwell Publisher,
108 Cowley Road, Oxford OX4 1JF, England

A free specimen copy is available on request, which should be made to the address above.

Basil Blackwell Publisher · Oxford · England

defining new mathematical properties or new functions are both sound, then there can be no objection to employing both of them, and considering both the entities introduced by one method and those introduced by the other. Since the 'possible worlds' on this interpretation are not directly characterized in epistemic terms, the appeal to epistemic possibility is not obviously relevant. All sorts of schemas may 'for all we know' turn out to be proper ways of defining functions or properties, just as all sorts of sets of equations may 'for all we know' define recursive functions, but on the proposed interpretation only those schemas which *are* proper modes of definition count as generating 'possible worlds', and of them we know that any number can be used side by side to generate a 'possible world' containing all the entities defined by any of them.

The model-theoretic result is that Kripke-models in which the possible worlds form a directed set under the accessibility relation validate a logic which is stronger than intuitionistic logic, and which may be characterized by adding to intuitionistic logic, as a new axiom-scheme, 'Either *Not-P* or *Not-Not-P*'. It is tempting to infer that neither Kripke's interpretation in terms of the possibility of acquiring new information nor the interpretation in terms of the enlargeability of the universe of mathematical objects would be acceptable to an intuitionist or quasi-intuitionist who did not accept this axiom.

To give in to this temptation, however, would be to misinterpret the significance of the model-theoretic result. The proof that the additional axiom is validated by arbitrary directed Kripke-models is non-constructive. Reasoning intuitionistically, we can specify a Kripke-model (with a decidable set of 'worlds', a decidable accessibility relation, and in which it is decidable whether or not an atomic formula is verified at a given 'world') in which the accessibility relation is directed (indeed, linear) but in which the axiom does not hold. Let the 'worlds' be the natural numbers, let greater numbers be accessible from smaller, and let the atomic formula '*P*' be verified at n just in case there is an m less than n such that the m-th digit in the decimal expansion of π begins a run of seven consecutive sevens in that expansion. Since we cannot, intuitionistically, assert that either there is no run of seven consecutive sevens in the decimal expansion of π or it is false that there is no such run, we can similarly not assert that the formula 'Either *Not-P* or *Not-Not-P*' holds in this Kripke-model.

Mathematically the point is trivial. Both intuitionistic logic and the logic with the additional axiom scheme have what is called the *finite model property*, and so are characterized by the classes of Kripke-models and directed Kripke-models, respectively, with finite sets of worlds. Since it can be proven intuitionistically that the added axiom holds in every directed Kripke-model with a finite set of worlds (assuming decidability of the set of worlds, the accessibility relation, and the verification relation between atomic formulas and worlds), our point is largely irrelevant to the mathematical analysis of the propositional logics. What it does show is that a philosophical position that rejects as invalid the axiom-scheme cited and also adopts one of the intuitive interpretations of Kripke's model theory described above as the skeleton of its semantic account of the meaning of its logical constants is not necessarily incoherent.

On both interpretations it is plausible to think of the class of 'possible worlds' as infinite. Thinking of them as 'evidential situations', and assuming that what can be established at a given moment of time may be represented as a formal system, it in effect follows from Gödel's incompleteness theorem that the class of 'possible worlds' required for a semantic account of mathematical language cannot be finite. Any finite set of formal systems, after all, can be combined in a single system, but for any one formal system, there are some mathematical truths not derivable within the system. Thus, given any finite set of 'evidential situations', there must be some 'evidential situation' in which information is available *beyond* the sum of the information available in the set. (Such an interpretation, with the 'possible worlds' corresponding to axiomatic systems, one being accessible from another if it is an extension, perhaps by the addition of a Gödel sentence as a new axiom, of it, bears some resemblance to the provability interpretation of S4 described by Kripke at the end of his 'Semantical Considerations on Modal Logic' (*Acta Philosophica Fennica* vol. 16, 1963).)

In the interpretation on which the 'possible worlds' correspond to different universes of properties and functions, if it is assumed that the properties and functions 'existing' at a given world are those which can be defined in some formalized language, the semantic paradoxes can be used to show both that the class of 'possible worlds' is infinite and that classical logic is inadmissible in reasoning that generalizes about the whole class. Any finite set of formal languages can be combined into a single language in which anything definable in any language of the set may be defined, but there can be no one language in which everything definable at all may be defined. (In its full generality, this thesis inherits the vagueness of the notions of a language and of definition; at least as far as *classical* languages are concerned, it may be taken to have been established by Tarski's proof that in any such language either some syntactical relations among its own expressions or the property of being one of its own true sentences must be indefinable.) The second claim, about the inadmissibility of classical reasoning about the whole system of 'possible worlds', is most easily established if we first accept the Platonistic claim that there are ordinal numbers so large that not all of them can be defined in languages beings like us could, even in principle, learn. Given this assumption, classical logic would lead us into the paradox of the least indefinable ordinal, for in our considerations of the system of 'possible worlds' we will certainly want to be able to speak of properties as being or not being defined at a given world. If we could do this in a classical language, we would be able to define the property of ordinals, 'being such that there is no property definable at any world which it is the least ordinal to possess'. This property would have to be defined at some world, and, given the Platonistic assumption that there are more ordinals than possible definitions, there would have to be a least ordinal possessing it, which is impossible. On the contrary assumption that every ordinal is definable, a version of Burali-Forti's paradox threatens.

The interpretation of Kripke's model theory that identifies 'possible worlds' with universes of definable functions and properties may not provide a foundation for strict intuitionism: it combines too happily with Platonistic set theories, and does not seem by itself to motivate the rejection of classical logic within, say, number theory. It does, however, seem to provide a coherent foundation for a quasi-intuitionistic position that allows only intuitionistic reasoning in some areas, and it would be a mistake to assume that the directedness of its accessibility relation would force its proponents to accept a logic stronger than the intuitionistic one.

460 Riverside Drive
New York, NY 10027, U.S.A.

© ALLEN HAZEN 1982

SETS WHOSE MEMBERS MIGHT NOT EXIST

By THOMAS BALDWIN

IN the course of the development of his essentialism Wiggins suggests that it is essential to a set such as {Eiffel Tower, Crystal Palace} that it have the members it does have, the Eiffel Tower and the Crystal Palace (*Sameness and Substance*, henceforth *S&S*, p. 114). Wiggins, I think, makes this suggestion primarily to provide an uncontentious example of an essentialist truth, although the suggestion is also directed *ad hominem* against Quine. For myself, I am in broad agreement with Wiggins on the issue of essentialism; but this particular suggestion is more contentious than he implies and merits attention in its own right.

Expressed in his favoured notation (cf. *S&S* pp. 108ff) Wiggins' claim is of the form:

(1) $\mathrm{Nec}(\lambda x)(a \in x \;\&\; b \in x)\langle\{a, b\}\rangle$

Some simplification is possible, however, for where a and b are both things whose existence is contingent, the same issues are raised by simpler claims of the form:

(2) $\text{Nec}(\lambda x)(a \in x)\langle\{a, b\}\rangle$

One way to dispute such a claim is by reference to the standard set-theoretic identity

(3) $\{a, b\} = \{x : x = a \vee x = b\}$

On the assumption that (3) is to apply to all possible worlds, a possible world in which a does not exist will be one in which $\{a, b\} = \{b\}$, and hence one in which $a \notin \{a, b\}$. Thus (2) will be false, in that there is a possible world in which {a, b} exists but does not have a as a member; or, in Wiggins' idioms (*S&S* pp. 115-6), given (3) it is possible to conceive of {a, b} without conceiving of it as having a as a member.

This argument relies on the standard set-theoretic identity (3), but in fact I think that rather than successfully controverting (2), it shows that there is something amiss with (3), at least as interpreted for the purposes of the argument. For, given (3), sets become mysterious things which change their identity from one possible world to another. Where a and b are distinct contingent objects, the truth-values of each of '$\{a, b\} = \{a\}$', '$\{a, b\} = \{b\}$', '$\{a, b\} = \emptyset$' will vary from possible world to possible world. In which case one's sense of what the set {a, b} really is crumbles. One way to see this is by noting that on this view sets conflict with the necessity of difference principle, which in Wiggins' essentialist idioms is

(4) $x \neq y \supset \text{Nec}(\lambda u)(\lambda v)(u \neq v)\langle x, y\rangle$

Furthermore, if one allows reference to merely possible objects (an issue I shall discuss later in this paper), so that with a merely possible but not actual object c one has $\{a, b, c\} = \{a, b\}$ in the actual world but not in all possible worlds, sets conflict with the necessity of identity principle:

(5) $x = y \supset \text{Nec}(\lambda u)(\lambda v)(u = v)\langle x, y\rangle$

An upholder of (3) might suggest at this point that all my argument so far has shown is that expressions of the form '{a, b}' are not rigid designators, even when 'a' and 'b' are. This, indeed, is certainly shown by the argument; nor is there anything novel in the thought that sets can be designated by non-rigid designators, for the point that set-abstracts are typically non-rigid was made years ago by Smullyan and is implicit in Russell's 'no class' treatment of them. However, what is disconcerting about accepting that expressions such as '{a, b}' are non-rigid, with rigid 'a' and 'b', is that we could have no better way of rigidly designating a set, and since sets are abstract objects, we surely need some such designators if we are to have an adequate grasp of them.

This attempt to refute (2) by appeal to (3) has failed, and it is now time to specify in more detail Wiggins' alternative view of sets. For Wiggins {a, b} does not exist where either a or b do not exist (*S&S* p. 113); Thus (3) is here replaced by

(6) $(\text{a exists} \;\&\; \text{b exists}) \supset \{a, b\} = \{x : x = a \vee x = b\}$

and the argument for (2) is immediate: if {a, b} exists, then so does a, and hence $a \in \{a, b\}$ wherever {a, b} exists. The 'converse' of (2)

(7) $\text{Nec}(\lambda y)(y \in \{a, b\})\langle a\rangle$

is, however, false on this view. For where b does not exist, neither does {a, b}; but b's non-existence does not entail a's non-existence, so it is possible for a to be, but not belong to {a, b}. Thus, on this view, the truth of (2) and falsehood of (7) exemplify nicely the kind of asymmetrical essentialist claims whose easy expressibility is such a strong point in favour of Wiggins' approach to essentialism.[1] Finally, returning to Wiggins' account of sets, for him '{a, b}' is a rigid designator wherever 'a' and 'b' are, unlike '$\{x : x = a \vee x = b\}$', and there is no problem in determining the identity of at least finite sets.

Yet Wiggins' approach raises a new problem. Sets are abstract objects and therefore, one would have thought, aloof from the vicissitudes of concrete objects. But since for Wiggins the existence of {a} is necessarily equivalent to that of a, it is hard to see how the contingencies of a's existence (where a's existence is contingent) do not affect the existence of {a}. When Paxton created the Crystal Palace, did he not, at least on Wiggins' view, equally bring into being {Crystal Palace}? Admittedly the existence of {Crystal Palace}, unlike that of the Crystal Palace itself, is presumably atemporal, and the destruction of the latter did not entail the non-existence of the former (or of {Crystal Palace, Eiffel Tower}, which is Wiggins' example of a set whose existence is assumed to be uncontentious). But this does not prevent there being, on Wiggins' view, apparently temporal and causal facts about the existence of {Crystal Palace} which are hard to reconcile with its status as an abstract object; e.g. that at a certain time Paxton, by doing something which he might not have done, brought it about that {Crystal Palace} exists. It may be objected that this consequence is only objectionable if one assumes a realist conception of abstract objects, and that if one replaces this assumption with a constructivist conception of abstract objects as products of human activity and thought, all is well. There is no space here to enter into the issues this constructivist view raises; so I propose to leave it aside with the thought that I do not know of any conclusive arguments for it, and, therefore, that it seems worthwhile to ask whether one can formulate an acceptable account of the existence of sets with members whose existence is contingent which is faithful to a more realist conception of abstract objects. A different objection will be that in my argument I have made the questionable assumption that necessarily equivalent propositions can be interchanged within temporal and causal contexts, an assumption which, it may be argued, has been shown to be dubious by the Frege-Church-Davidson arguments about extensionality. Again, there is no space here for the issues raised; I shall simply state without argument that I do not believe that my assumption is undermined by the Frege-Church-Davidson arguments. It is rather non-Russellian assumptions as to what is a singular term that are undermined by those arguments, though this point does not exhaust their significance.

---

[1] It is worth remarking that Moore had spotted the similar assymetry in the essentialist truths (internal relations) which apply to the part/whole relation; cf. 'External and Internal Relations' p. 288 in *Philosophical Studies.*

A familiar feature of much recent thought about modal concepts has been an analogy between possibilities and times, and I want now to suggest that by pursuing this analogy in the present case one can find an acceptable solution to these puzzles about sets whose members exist only contingently. Wiggins, as I have observed, implies that {a} exists even if a does not exist now, as long as it has existed (or, presumably, will exist). The analogy between time and modality suggests a further extrapolation: {a} exists even if a has never existed, nor will ever do so, as long as it might exist. Return now to the standard identity (3): in my initial discussion of (3) I assumed that values of 'x' were restricted to things which actually exist (hence, on that view, where a does not exist $\{a, b\} = \{b\}$). Suppose, however, that one takes the domain to include possible as well as actual things, and interprets (3) accordingly. On this interpretation, as long as a might exist, $a \in \{a, b\}$, and therefore, even if a does not exist $\{a, b\} \neq \{b\}$. In this way the problems which afflicted the original account of sets based upon (3) are now removed: since the condition on set-membership is now not sensitive to questions of actual existence or not, sets keep their members unchanged in all possible worlds. There is no puzzle about the identity of {a, b}, nor any threat that it will violate principles of the necessity of identity and difference ((4) and (5)).

This approach also keeps clear of the difficulty which Wiggins' approach faces. Since {Crystal Palace} exists provided that the Crystal Palace is a possible object, its existence is not necessarily equivalent to the existence at some time of the Crystal Palace itself, and stands in no temporal or causal relation to Paxton's creation of the latter. Perhaps we could not conceive of {Crystal Palace} without Paxton's actual creation of the Crystal Palace, but once one separates conditions for the existence of a set from conditions of its conceivability by us, this is no objection. A further feature of this approach is that despite its affirmation of (3), suitably interpreted, it also preserves Wiggins' essentialist intuition (2). For on this approach '$a \in \{a, b\}$' does not entail 'a exists'; thus it is not possible for {a, b} to be without having a as a member. What this approach does conflict with, however, is Wiggins' further intuition that the 'converse' of (2), (7), is false: for since {a, b} exists whether or not b actually exists, it is not possible for a to be without belonging to {a, b}.

One way to compare my approach with that of Wiggins is to view them as different responses to the tensions which the standard account of (3) reveals between the abstractness of sets and the contingency of some of their members. Wiggins, in effect, levels sets down to the contingencies of their members. My approach cuts the membership condition for sets free from the contingencies of actual existence and thereby raises it to the same level of abstraction as that of sets. Wiggins' approach seems to me incompatible with the notion of a set as an abstract object, at least on a realist account of abstract objects; the question for my approach is whether it suffers from the opposite fault of failing to do justice to the realities of actual non-existence. Obviously there is no problem about introducing into set-abstracts an extra condition restricting the domain to what actually exists. But that does not remove my commitment to the actual existence of sets some of whose members only possibly exist.

My defence of this claim starts from reflecting on what is to be said about the identity of a set such as {Crystal Palace, Eiffel Tower} under the supposition that though the Eiffel Tower exists, the Crystal Palace has never done so, and never will do so. Once one rejects Wiggins' view that the set does not exist, and the view dictated by (3) as initially interpreted, that the set is just {Eiffel Tower}, there seems no rational alternative to the conclusion that, under this supposition, in these possible worlds, the set just is {Crystal Palace, Eiffel Tower}. But if in these possible worlds there is a set one of whose members does not exist in these worlds, it seems arbitrary to suppose that the same should not apply to the actual world, i.e. that there should not actually be sets some of whose members do not actually exist. Indeed if one rules this out, one seems bound to fall back on the view that if one were to create something which, in fact, never has existed and never will do so, one would thereby bring it about that sets would exist which do not actually do so. Yet this is just a hypothetical variant of the consequence which I find so unpalatable in Wiggins' account.

I do not deny that there are difficulties in the conception of merely possible objects. In particular, the Barcan formula and some naive model-theoretic talk of domains of possible objects may encourage a suggestion which must now be repudiated: that it is somehow possible for us to ascend in thought from the actual world to a point of view, the modal analogue of the divine view of the *species aeternitatis*, from which all possibilities are equally accessible to us. In place of this what is required is a means of specifying uniquely a merely possible object, without requiring any supernatural skills. Consider, then, the chocolate cake I might have made yesterday. I did not make any such cake but I might have done, and in a world in which I did {Crystal Palace, Choc Cake} ≠ {Crystal Palace}. Preserving this difference in the actual world requires that {Crystal Palace, Choc Cake} actually exists with Choc Cake as a member, even though Choc Cake does not actually exist. But what is Choc Cake? There are many possible worlds in which I made just one chocolate cake yesterday; within which, one may ask, did I make Choc Cake? Since non-actual possible worlds cannot in principle be uniquely specified, this question cannot be answered. But it will be sufficient if one can specify a class of possible worlds such that the differences between these possible worlds would not affect the identity of a chocolate cake made in them. Thus consider worlds just like the actual world except that instead of reading the newspaper yesterday afternoon, I made just one chocolate cake, following my favourite recipe and using only the ingredients most readily available to me. These worlds will differ with respect to trivial details, such as the precise time at which I switched the oven on; but details such as this will not affect the identity of a cake. More seriously, the worlds will also differ with respect to the precise grains of flour which go into the cake, and this point raises the question as to whether one should say of an actual cake that it would not be the cake that it is if it was not made of precisely the grains of flour that it was made of. This question is comparable to familiar questions about the identity of Theseus' ship through repairs to its woodwork (cf. *S&S* pp. 92ff), and a plausible assumption is that identity of a cake does not require exact identity of its constituents; so that as long as in each world the cake is made from nearly the same ingredients, in each world the same cake is made. But it should be noted that even if this assumption is rejected, Choc Cake can still be uniquely specified, by actually identifying precise quantities of flour etc., and then restricting possible worlds in which Choc Cake is made to worlds in which just these quantities are used.

I conclude, therefore, that Choc Cake can be uniquely specified; and thus that the set {Crystal Palace, Choc Cake} is a determinate object of thought, whose actual existence is compatible with Choc Cake's merely possible existence.

University of York,
Heslington, York YO1 5DD

© THOMAS BALDWIN 1982

INTERNAL STRUCTURE AND ESSENCE

By A. C. GRAYLING

'TIGERS', says Kripke, 'cannot be defined simply in terms of their appearances; it is possible that there should have been a different species with all the external appearances of tigers but which had a different internal structure and therefore was not the species of tigers.' (*Naming and Necessity* p. 156; cf. also pp. 119-121.)

Does the notion of internal structure as essence make sense? It does not. Consider the fact that internal structures are nested. Tigers are composed of a certain arrangement of internal organs and other physiological structures, which in turn have an internal structure of cells, which in turn have an internal structure of molecules... atoms... subatomic particles... and so on, perhaps, to quanta of energy (and beyond?). Which level of internal structure is *the* level constituting the essence? Suppose two tigers have the same internal structure at all levels 1 to *n* but differ thereafter. Which is the tiger?

It might be replied that any level of internal structure entails that all other levels are uniformly correspondent, so that specifying similarities at one level is sufficient. But if levels of structure are uniformly correspondent, then there is every reason to suppose that the correspondence runs up to and includes external structure, so that – a point Kripke denies – external structure (which is to say, external appearance) is enough to settle that something is a thing of a certain kind.

Moreover, if differences of internal structure mark differences of kind, then a counterintuitive result follows, to the effect that each individual is its own kind. For, every individual is unique – this tiger has longer hairs on its left ear than that one – and such differences result from differences in internal structure. This unpalatable result can be blocked by saying that it is a logical sum of the internal properties which count, namely the logical sum of inclusive disjunction. However, this notion is one which Kripke contests in connection with external appearance; yet here the appeal to a logical sum of properties arises in connection with internal structure, which is, epistemologically speaking, a far less appetizing situation than relying upon a logical sum of appearances for kind differentiation.

Internal structure is invoked principally because it seems precise and unequivocal to give the essence of, say, gold or water, by saying that all and only the stuff of atomic number 79 in any world is gold, and water is all and only the stuff of molecular structure H_2O in any world. At first blush elements and compounds like these may seem to require an alternative analysis, given their simplicity as against the complexity of biological kinds; but the same arguments go through. H_2O is molecular stuff with an 'internal' structure of atoms which in turn have an internal structure...and so on.

It is logically possible that internal structures be infinitely nested. Because this is so, it is not open to essentialists to say that *whatever* internal structure something x may have – that is, even if we do not know or cannot determine what the internal structure of x is – nevertheless x belongs to a kind K in virtue of it. (Kripke does this: cf. ibid. pp. 120–1.) For, possession of publicly applicable criteria for sorting things into kinds is a necessary condition of sorting things into kinds (and *ipso facto* of kind terms being publicly usable). We *do* sort things into kinds; it follows by *modus ponens* that we possess such criteria. But if it is logically possible that internal structures be infinitely nested, and if we have no way of choosing *the* level of internal structure which is essential to the kind, then such criteria can have nothing to do with internal structure: if we could not in principle know what makes x and K, we could not recognise some further thing y as a K owing to its (putatively) having the same internal structure as x ('putatively' because the internal structure is recognition-transcendent – it is '*whatever* it is'). Thus if something x is a K in virtue of its essence, then essence and internal structure cannot be identified with each other.

It is worth remarking, by way of codicil, that if, as seems to be the case, recent arguments against origin as essence work (e.g. M. Slote, *Metaphysics and Essence*, Oxford 1974, pp. 6–7, P. Johnson, 'Origin and Necessity', *Philosophical Studies* 32, 1977, pp. 413ff), then the result of the conjunction of those arguments with this one appears to be: either we rest content with a notion of nominal essence, or – and perhaps this comes to the same thing – we eschew talk of essence altogether.

Magdalen College, Oxford © A. C. GRAYLING 1982

PREDICATIVE AND CONSTRUCTIVE KNOWLEDGE

By D. E. OVER

MICHAEL DUMMETT has distinguished *propositional* and *predicative* knowledge in his paper, 'Frege's Distinction between Sense and Reference' (*Truth and Other Enigmas*, Duckworth, 1978). His example of predicative knowledge (on p. 125) is:

(1) The police know, of Redmayne, that he murdered Bexley.

As an example of propositional knowledge, he gives (again on p. 125):

(2) The police know that the man who hid on the roof murdered Bexley.

When (1) is true Dummett holds that the police know who murdered Bexley. But he denies that this follows from (2), unless there is an independent reason for saying that the police know who hid on the roof.

Dummett seems to be introducing new terminology for the well known distinction between *de dicto* (propositional) and *de re* (predicative) knowledge, a distinction which is generally extended to the other propositional attitudes. But Dummett's main purpose is to analyse Frege's argument for distinguishing sense and reference, and this argument, according to Dummett, is based on two premisses, one of which is relevant here:

(3) Every piece of predicative knowledge rests on some piece of propositional knowledge. (p. 126).

Frege held not only that the concept of sense must be added to that of reference in semantics, but also that no word could have a reference without a sense. It follows from his position, as interpreted by Dummett, that no one could know the reference of a word without assigning a sense to it (no one could have what Dummett calls bare knowledge of the reference of the word). To use a metaphor developed by Dummett, the sense of a word is the route to its reference, and everyone who associates a reference with a word must have some route to the reference, some way of identifying the reference.

In support of Frege's position Dummett would use (3) in the following way. Some philosophers might argue that, if (1) is true, then the police have bare knowledge of the reference of 'Redmayne', i.e. they know that 'Redmayne' refers to a certain man whom they can find or pick out without using anything that could be called a Fregean sense of the name. Given (3) Dummett would reply that the police's predicative knowledge referred to in (1) must be based on some propositional knowledge, perhaps that referred to in:

(4) The police know that the mayor of the city murdered Bexley.

Intuitively, it seems right to conclude that (4) supports (1) if Redmayne is the mayor. Certain other descriptions would support (1) equally well if they were substituted in (4), e.g. 'the man in the cells' or 'the tallest man in the identity parade'. Whatever description of this type is used, Dummett can *try* to argue on Frege's behalf that its role in (4), as the basis of (1), is to express the sense of 'Redmayne' for the police, the way the police have of finding the reference of 'Redmayne'.

The concern of this paper is not with the persuasiveness of Dummett's analysis of Frege's position on sense and reference. Our topic will be (3) and its justification. The fundamental problem about (3) for Dummett is to explain why some examples of propositional knowledge support predicative knowledge, while others do not. Assuming that Redmayne is both the mayor and the man who hid on the roof, why do we feel that (4) supports (1) much more strongly than (2) does on its own? Why is it that

(5) The police know that the man who murdered Bexley murdered Bexley

does not seem to support (1) at all, even if Redmayne is the murderer? Clearly, this problem, in other terminology and with a different perspective, is a special case of the problem of accounting for any *de re* context in terms of a *de dicto* context. But although Dummett endorses (3), he does not try to justify it by answering the above questions.

Our purpose will be to indicate how Dummett could tackle these questions by making use of a distinction central to his own work in semantics, that between constructive and non-constructive knowledge. (The importance of this distinction is stressed again and again in many of the papers in *Truth and other Enigmas*.) The concept of the constructive is at its clearest in mathematics, in which the basic motion is that of a constructive proof.[1] A mathematician could be said to have constructive knowledge if he could justify his position by means of a constructive proof; if, however, he could only justify his position by means of a non-constructive proof, then he could be said to have non-constructive knowledge. Consider:

(6) There is a number which is a sum of two cubes in two different ways.

There are two ways in which a mathematician – say, Hardy – can know that (6) is true. First, there is the *de re* or predicative interpretation of his knowledge:

(7) There is a number which is such that Hardy knows that it is a sum of two cubes in two different ways.

Second, the *de dicto* or propositional case:

(8) Hardy knows that (there is a number which is a sum of two cubes in two different ways).

Corresponding to (7) and (8) there are two different ways in which Hardy could prove (6). He might actually present a particular number and show that it had the property referred to in (6). He would then have a constructive proof of (6) and (7) would clearly follow. Hardy would know, of a certain number, say 1729, that it was a sum of two cubes in two different ways. Alternatively, Hardy could have a non-constructive proof of (6); he might prove (6), for example, simply by assuming its negation and deriving a contradiction. He might then be unable to say which particular number had the property in question, and so (8) would be true but not (7). Using these examples, we can say that Hardy has constructive knowledge in (7) and non-constructive knowledge in (8).

---

[1] For the concept of a constructive proof, see Michael Dummett. *Elements of Intuitionism.* (O.U.P., 1977), Ch. 1.

Constructive knowledge could be taken as a wide notion which is not necessarily based on constructive proofs. Suppose that Hardy only hears of the truth of (6) from Ramanujan. Hardy does not have a proof in this case, but in the ordinary sense he has knowledge, since he recognizes that Ramanujan is a mathematician of genius and may be accepted as an authority. Even here there is a distinction between (7) and (8). If Ramanujan tells Hardy that 1729 is the relevant number, an example of a number which is a sum of two cubes in different ways, then (7) is true. Hardy can specify a particular number which makes (6) true; he has constructive knowledge. But if Ramanujan does not give Hardy an example of a number with the required property, then only (8) is true, and Hardy has non-constructive knowledge.

Dummett wishes to extend and generalize the concept of the constructive; he wants to take it from mathematics and modify it so that it can be applied to natural language. He replaces the concept of proof with that of an effective *procedure* in a wide sense.[2]

Following Dummett, we would say that to have constructive knowledge that a disjunction is the case is to have an effective procedure that will establish in a finite time which disjunct is the case. To have constructive knowledge that an existential sentence is the case is to have an effective procedure that will pick out in a finite time a particular object which satisfies the predicate. Take:

(9) Someone in the identity parade murdered Bexley.

The police may have only general, non-constructive grounds for knowing that (9) is true. (The murderer must be one of four people who were at the scene of the crime. All of these are in the identity parade.) So the police will be unable to pick the person who satisfies 'murdered Bexley'. If (9) is replaced by the corresponding disjunction, containing the names of the people, in the parade, then the police will be unable to say which disjunct is true. The knowledge of the police is propositional or *de dicto* in this example.

---

[2] On Dummett's wide notion of an effective procedure, see his paper, 'What is a Theory of Meaning (II)?', in *Truth and Meaning*, ed. by G. Evans and J. McDowell. (O.U.P., 1976), especially p. 81.

Suppose, however, that the police have constructive knowledge that (9) is true. They may be able to pick out the murderer straightaway; and they would then certainly have predicative or *de re* knowledge. They must at least have a procedure that leads to the murderer in a finite time, such as checking the finger prints of all the suspects in the parade. Again, they certainly have predicative knowledge if they have followed the procedure to its end and pointed out the guilty party. But suppose they have the procedure but do not follow it? They know that the murderer is in the identity parade, that they have his finger prints, and that they could, with a little trouble, separate him from the rest; but for some reason they do not do this. Do they have predicative knowledge that (9) is true? That is, do they know of someone in the identity parade, that he murdered Bexley? It does seem natural to say that they do, although this may involve some extension of our admittedly vague understanding of predicative knowledge. The police in this case are in a much stronger position than someone who simply has *de dicto* or propositional knowledge that (9) is true; the police are much closer to central cases of predicative or *de re* knowledge.

Returning to earlier examples, we can now explain our intuitions about (2) and (4). Without more background information, it seems probable that the police have only non-constructive knowledge in (2). It seems probable that they do not know how to find effectively the man who hid on the roof; clues tell them that he was there but not how to pick him out from the population at large in an effective way. On the other hand, (4) suggests that the police do have constructive knowledge. We assume that they have an effective way of finding whom 'the mayor' refers to. Of course, this does not mean that they can necessarily arrest him; perhaps he has already left the country. But the police, we assume without more information, can say which member of the city is the mayor.

In (5) our feeling that this is non-constructive knowledge is at its strongest. We apparently assume that the police have inferred that just one man murdered Bexley, that they know that 'the man who murdered Bexley' refers to some man, whom they cannot pick out or separate effectively from the rest of the population. We take this to be the case, unless we are told that the police do have constructive grounds for their knowledge in (5). Without a background story, we naturally imagine that (5) is a prime example of non-constructive knowledge, which can only be *de dicto* or propositional.

As we noted earlier Dummett denies that (1) follows from (2), except in special circumstances, which we can now see are those in which (2) refers to constructive knowledge. Without hearing about special circumstances it is reasonable to interpret (2) as being about non-constructive knowledge. On the other hand, it is reasonable, without a special story, to take (4) as being about constructive knowledge; and consequently (4), and not (2), supports (1). (Here, of course, we continue to assume that Redmayne does happen to be both the man who hid on the roof and the mayor.)

Noting the connection between constructive knowledge and predicative knowledge leaves many deep questions concerning these concepts unanswered. What exactly is an effective procedure outside of mathematics? How exactly do we use procedures to pick out individuals? What kind of names or indexicals can be at the end of these procedures? What are the objects referred to by these names and indexicals? Some light may be thrown on such questions by studying the interrelation between the constructive and the predicative; but our present concern is with (3). Dummett endorses (3), saying, 'I do not believe that there is any way to explain predicative knowledge save in terms of propositional knowledge' (p. 126). He seems to imply that propositional knowledge is more fundamental than predicative knowledge; and we can now see that this position needs to be qualified from Dummett's point of view.

Dummett argues strongly (in many of the papers in *Truth and Other Enigmas*) that non-constructive theories of meaning stand in need of justification. How is it possible, he asks, for us to have non-constructive knowledge of a mathematical or physical fact? How can we form a notion of such a fact? He points out that we first learn to make statements on a constructive basis. For instance, we would be taught to assert (6) when we knew of a particular number which was a sum of two cubes in two different ways, or at least when we knew of an effective way of finding such a number. We would generally learn what Dummett calls assertibility conditions for sentences; and we could display our linguistic understanding in our assertions. Dummett demands to know how, on this constructive basis, we are justified in taking the leap to non-constructive principles of reasoning, which can only be justified in terms of an objective concept of truth conditions. He wants to know how we grasp this concept and display our understanding of it. Now we could summarize Dummett's position by describing it as holding that the constructive is more fundamental than the non-constructive in semantics; and it would follow, from the relation between the constructive and the predicative, that predicative knowledge is of fundamental importance in semantics.

Thus principle (3) must be carefully qualified; it must not be taken to imply that all propositional knowledge is more basic than any predicative knowledge. The point of (3) is to state that all predicative knowledge must be founded on certain propositions which, *it could be argued*, have Fregean senses. It is notoriously difficult to give a general characterization of the terms which may occur in these propositions; some proper names in certain circumstances, some use of indexicals, and certain definite descriptions in some cases ('the mayor' but not 'the man who hid on the roof') would qualify.[3] But it is very clear in many examples how predicative knowledge rests on *constructive* propositional knowledge. It is easy to see that (1) would be established if the police could point to Redmayne in an identify parade and say, 'That man murdered Bexley.' The police would have constructive propositional knowledge that that man murdered Bexley, and (1) could rest on this basis.

Thus (3) should read:

(3.1) Every piece of predicative knowledge rests on some piece of *constructive* propositional knowledge.

Propositional knowledge is not in general more basic than predicative knowledge. Some propositional knowledge is non-constructive, and this, from Dummett's point of view, means that it stands in need of justification not required for constructive knowledge and the associated predicative knowledge.[4]

Sunderland Polytechnic, Chester Road,
Sunderland SR1 3SD

© D. E. OVER 1982

---

[3] Kaplan tried to characterize these terms as 'vivid names'. See 'Quantifying In', *Words and Objections*, ed. by D. Davidson and J. Hintikka (Reidel, 1969). Tyler Burge, in *'Belief De Re', Journal of Philosophy*, Vol. 74 (1977), pp. 338-63, comments on Kaplan and concludes that people must have *de re* attitudes before they can be said to understand a language. Burge's position follows if, as argued in this paper, *de re* and constructive knowledge are closely related.

[4] I should like to thank Dorothy Edgington for helpful discussions on the topic of this paper.

GOLDMAN ON WHAT JUSTIFIES BELIEF

By DONALD E. STAHL

IN his recent article, 'What is Justified Belief?' Alvin Goldman says that he wants to 'specify when a belief is justified,' (in George S. Pappas, ed., *Justification and Knowledge*, Boston, D. Reidel, 1979, p. 1). This seems to me to mean 'give necessary and sufficient conditions.' What he does, however, is to provide a putatively sufficient condition, his

> (10) If S's belief in p at t results from a reliable process, and there is no reliable or conditionally reliable process available to S which, had it been used by S in addition to the process actually used, would have resulted in S's not believing p at t, then S's belief in p at t is justified. (op. cit., p. 20.)

Reliability 'consists in the tendency of a process to produce beliefs that are true rather than false,' (p. 10) and 'A process is conditionally reliable when a sufficient proportion of its output beliefs are true *given that its input beliefs are true*,' (p. 13, original italics).

Although 'We can and do regard certain beliefs as more justified than others,' Goldman intends to 'treat it [justification] here as categorical in the interest of simplicity,' (p. 10; Goldman apparently believes that 'categorical' is the antonym of 'relative'). The belief-forming processes concerned are restricted by Goldman 'to *cognitive* events, i.e., events within the organism's nervous system,' (p. 12). They are types rather than tokens, and Goldman recognizes 'A critical problem concerning... the degree of generality of the process-types,' since

> the degree of generality will partly determine the degree of reliability. A process-type might be selected so narrowly that only one instance of it ever occurs, and hence the type is either completely reliable or completely unreliable. (This assumes that reliability is a function of *actual* frequency only.) If such narrow process-types were selected, beliefs that are intuitively unjustified might be said to result from perfectly reliable processes; and beliefs that are intuitively justified might be said [to] result from perfectly unreliable processes. (p. 12).

Goldman thinks that it is a matter of our choosing 'ordinary' process-types. If we do this, he thinks, such unwanted results will not occur; but the problem consists in the fact that we do not have a general account of this sort of ordinariness.

There is also a closely related difficulty Goldman does not notice. Since what causes a belief is a process-token of a certain process-type, the same process (-token) may instantiate different process-types, and some of these types may be reliable, while others are not. Goldman's idea is that F-processes are reliable, and therefore produce justified beliefs, while G-processes are not and produce unjustified beliefs. But what are we to say of the belief produced by a process which is both F and G?

Let us assume that 'reasoning' (one of Goldman's own examples) is a reliable process; i.e., that reasoning results in more true beliefs than false ones; and let us assume that 'affirming the consequent' is a species of reasoning, and that it results in more false beliefs than true ones. According to (10), then, the belief resulting from a process of affirming the consequent is justified (assuming that there is no other reliable or conditionally reliable process available to *S*, etc.) since it results from a reliable process, viz. reasoning; and also unjustified, since it results from an unreliable process, viz. affirming the consequent – or so it seems. Goldman is actually uncommitted to the result that a belief resulting from an unreliable process is unjustified, since he gives only a sufficient condition for justifiedness. Further, Goldman might stipulate that it is the reliability or unreliability of a process as described in the most specific or determinate relevant way which confers or fails to confer justification.

However, is not the following case possible on Goldman's principles? *S*'s belief that *p* is the result of a process which qualifies as reliable by Goldman's criterion; it produces true beliefs more often than false ones – but just barely. *S*'s belief is therefore justified according to Goldman. Since it is a token, it may be of more than one type, though, and in this case it is; it's also of another type, which is *very strongly* unreliable. This type almost always produces false beliefs. Knowing this about the process, would not our intuitions tell us, *contra* Goldman, that the belief in question was not justified? Nor can specificity or determinateness be counted on to provide a correction, for the types involved may simply be unrelated to each other with regard to specificity or determinateness. Reasoning is neither more specific nor more determinate than memory (another of Goldman's examples), and *vice versa.*

In case anyone should object that it is implausible to suppose that the same process could be both an instance of memory and an instance of reasoning, I must point out that psychologists do not find it so. On the contrary, in fact. Consider the following passage from Mary J. Naus and Frank G. Halasz' 'Developmental Perspectives on Cognitive Processing and Semantic Memory Structure', in Laird S. Cermak and Fergus I. M. Craik, eds, *Levels of Processing in Human Memory*, Millsdale, New Jersey, Lawrence Erlbaum, 1979, p. 271.

> These [eight preceding] studies of constructive memory are based upon a view in which rather than being considered simply a snapshot of events filed in memory at storage and faithfully retrieved at recall, memory is thought of as a construction of an internal interpretation of the stimulus by the subject. This cognitive construction consists of selecting relevant information from incoming stimuli as well as elaborating this information on the basis of stored semantic knowledge. Similarly, the process of retrieval from memory is seen as an active reconstruction process in which the remembered material is constructed from information available in semantic memory and in the recall environment. In Bartlett's (1932) terms, memory involves 'going beyond the information given' and actively embellishing stimulus information. In the present view, this elaboration and integration of environmental stimuli in conjunction with semantic memory is viewed as an automatic process.

An even worse defect of Goldman's account remains to be discussed. There seem to be no belief-producing processes of the kind Goldman requires. Consider some of the beliefs *S* has at *t*, which he has never thought of, and will never think of; such as that tweed coats are not made of salt, and that zebras in the wild do not wear overcoats. (See Daniel C. Dennett, *Brainstorms*, Montgomery, Vermont, Bradford Books, 1978, Chapters 3, 6 and 16.) Obviously, most of *S*'s beliefs are like this, and while it is justified beliefs which are most obviously attributable to *S*, it is also the case that we sometimes, in explaining his more inept actions, attribute to him unjustified beliefs which have never occurred to him – very much as I attributed to Goldman the unjustified belief that 'categorical' is the antonym of 'relative'. What are the processes which have produced these unthought-of beliefs in *S*? I submit that the only plausible candidate for such a process is simply *S*'s total cognitive process up to *t*. This means that all such beliefs of *S* are either all justified or all unjustified according to Goldman, since they all have precisely the same causal ancestry. There are no *differing* causal processes, some reliable and some not, to account for them. Goldman's account of justification seems appropriate not to beliefs in general, but only to what Dennett (Chapter 16) calls opinions.

University of Illinois at Chicago Circle,
Box 4348, Chicago, Illinois 60680, U.S.A.

© DONALD E. STAHL 1982

MANIPULABILITY THEORY AND EVENT TYPES

By DOUGLAS EHRING

VARIOUS philosophers who have rejected Humean and Neo-Humean accounts of causation have adopted manipulability theories of causation. Manipulability theories analyse sentences of the form 'a causes b' roughly as follows: we can bring about (or prevent) b by producing (or preventing) a, but we cannot bring about (or prevent) a by producing (or preventing) b. A standard objection to manipulability theories is that many events which are causally connected are either in practice or in principle (in part due to laws of nature) not within the influence of human action. One line of reply to this criticism is taken by Gasking:

> And one can sometimes properly say of some particular happening, A, that it caused some other particular event, B, even when no one could have produced A, by manipulation, as a means of producing B. For example, one may say that the rise in mean sea-level at a certain geological epoch was due to the melting of the Polar ice-cap. But when one can properly say this sort of thing it is always the case that people can produce events of the first sort as a means of producing events of the second sort. (Douglas Gasking, 'Causation and Recipes', p. 129 in T. Beauchamp, ed., *Philosophical Problems of Causation*, Dickenson 1974).

This line of modification will not, however, provide an adequate rejoinder to this criticism. The following counterexamples are meant to show that this analysis in terms of 'sorts of events' is consistent neither with our notion of 'causally connected' nor with that of 'causally prior'.

(1) Suppose that a and b are non-manipulable events. In addition, c, d, e and f are manipulable events where c and e are of the same relevant sort as a, and d and f are the same sort of events as b. We can suppose that d can be brought about by producing c, but that c cannot be brought about by producing d. But suppose that e cannot be brought about by f and f cannot be brought about by e. The modification suggested by Gasking requires us to assert that a causes b since a and c are the same sort of events and b and d are the same sort of events, and we can bring about d by means of c but not c by way of d. In fact, however, a and b may fail to be causally connected just as e and f are causally unconnected. Since a and b are respectively the same sort of events as e and f, where e and f are causally unconnected according to a straightforward application of manipulability theory, we cannot exclude the possibility that a and b also are causally unconnected.

(2) Objection (1) may be changed slightly to show that Gasking's refinement leaves us with a difficulty in accounting for causal asymmetry. Assume that the same events are at stake as in objection (1) and that they are instances of the same types of events. In turn suppose again that we can bring about d by producing c, but that we cannot bring about c by producing d. In place of our assumption about e and f, assume that we can bring about e by producing f, but that we cannot bring about f by producing e. On the basis of Gasking's rejoinder we must conclude that a causes b and b causes a.

(3) A similar difficulty to (1) is the following. Instead of two types of events we may use *four* types. Again a and b are non-manipulable events. Now suppose that c, d, e and f are manipulable events. a and c are the same sort, as are a and e, but c and e are of different sorts (the 'of the same sort' relation is not transitive). Similarly, b and d are of the same sort, as are b and f, but d and f are not of the same sort. Suppose that we can bring about d by producing c and we cannot bring about c by producing d. But now assume that we cannot bring about e by means of f or f by means of e. What are we to say about the possible causal relation between a and b? As Gasking's suggestion stands, we must conclude that a causes b since a and c are of the same sort, as are b and d, and c causes d by a straightforward application of manipulability theory. This, however, will not do. In fact, a and b may not be causally connected, just as e and f fail to be causally connected. This possibility must be excluded before we settle on the judgment that 'a causes b'.

(4) Assuming the same constellation of sorts of events as in objection (3), suppose that we can in fact bring about e by producing f, but that we cannot bring about f by producing e. In this case, given that we can bring about d by means of c, but not c by means of d, we must conclude on Gasking's analysis that a causes b and b causes a. Thus causal asymmetry has not been adequately treated.

These counterexamples involve the possibility that the modified theory might generate a mistaken answer to the question 'does a cause b?' where a and b are non-manipulable events. One line of reply to these cases is the following: it might be admitted that such problem cases are logically possible, but they will in fact never occur. The modified theory may be interpreted as including criteria for events being of the same type which will guarantee that the non-manipulable events are correctly classed with one pair of manipulable events rather than the other pair. In the case of objection (3) for example, we can suppose that there will be some feature k which is present in the case of e and f but not present in the case of c and d, where k's causal role can be established by a straightforward application of manipulability theory. k will either be compresent with a and b or it will not. If k is present with a and b then our criteria will determine that a and b are to be classed with e and f rather than c and d (and vice versa if k is not present with a and b). The failure of e to cause f is thus traceable to a difference in the circumstances that is determined to be relevant by a direct application of manipulability theory. We can then extrapolate to the non-manipulable case.

However, this reply (which I owe to an ANALYSIS referee) will not serve to guarantee that a and b will be correctly classed with the appropriate pair of manipulable events. The criteria determine that a and b are to be classed (or not classed) with e and f in part because of the presence (or absence) of k in both the manipulable and non-manipulable contexts. But whether or not k is relevant to the possible causal relation of two events will generally depend upon the presence of additional features of the circumstances. Thus, for example, the presence of oxygen is relevant to the possible causal relation between the striking of a match and the lighting of the match, only given the presence of other conditions such as the state of the striking surface. In the examples above, a and b are non-manipulable events, and they are so in virtue of some set of features $g_1, g_2, \ldots, g_n$ of the circumstances. Since $g_1, g_2, \ldots, g_n$ determine non-manipulability they fail to be present in the manipulable cases. k in the presence of $g_1, g_2, \ldots, g_n$ may or may not play a comparable causal role as k in the manipulable case. $g_1, g_2, \ldots, g_n$ may counteract, for example, k's interference with the causal relation between a and b, although k does prevent the causal relation between e and f in the manipulable case where $g_1, g_2, \ldots, g_n$ are absent. In turn, no direct application of manipulability theory can determine the causal relevance of k in the presence of $g_1, g_2, \ldots, g_n$ since the latter features would insure that no such manipulation was possible.

Southern Methodist University,
Dallas, Texas 75275, U.S.A.

© DOUGLAS EHRING 1982

REPLY TO LOWE ON ACTIONS

By JENNIFER HORNSBY

ONE issue about the relations between *action* and *causation* is this: How are '*A* caused *x*' and '*A*'s action caused *x*' related? There seems to be no actual entailment between these sentences in either direction. For, first, a man can cause things without acting. And, second, when we attribute effects to events (including actions) we are prepared to look further afield than we are when we attribute effects to agents.[1] Still, I think that we can say at least that an action causes many of the events that the agent causes and that a person causes many of the events that his action causes. So suppose that some event *x* is both caused by an agent *A* and caused by his action. *A* caused *x*, then, and it seems plausible that his causing *x* is his action. (Or so I argued in *Actions*, Routledge and Kegan Paul, 1980). But *A*'s action caused *x*. So we have it that *A*'s causing *x* caused *x*. Dr E. J. Lowe finds this conclusion 'distinctly odd'. See his '"All Actions Occur Inside the Body"', ANALYSIS 41.3 (June 1981) pp. 126-29.

Lowe quotes the following sentence from a footnote in Chapter One of *Actions*: '"A person caused something" can be taken to say that an event in which a person played a role caused something'. From this and the conclusion that I have reported he finds distinctly odd, he purports to derive a further conclusion – equally unacceptable to us both – that *y*'s (an event's) causing *x* either caused *x* or partially caused *x*. And he says of the view he is attacking that it 'arises from a serious misconception of the relationship between the notions of action and causation'.

So far as '*A*'s causing *x* caused *x*' is concerned, this is no odder than the plainly truistic '*e*'s cause caused *e*'. So far as the unacceptable conclusion is concerned, it does not follow by any known logical principle from the sentence Lowe quotes. That sentence was explicated in Chapter Seven of *Actions*, to which the footnote from which Lowe quotes refers; its purport is given again in my opening paragraph above. Lowe claimed that the quoted sentence committed me to the reducibility of agent causation: he supposed that it supplied a reductive recipe licensing the replacement of an agent's name by a name of an event in any causal sentence. But this is rather like supposing that, where 'The table is green' can be taken to say what colour the surface of the table is, we can replace 'the table' by 'the table's surface' whenever we ascribe properties to the table. Tables and their surfaces share some of their properties, but not all of their properties. Agents and their actions share some of their causal properties. It does not follow that they share all of their causal properties.

---

[1] Or so it appears: it is as if we cannot leave aside questions of responsibility when making attributions to persons. (Consider the examples in J. E. Atwell, 'The Accordion Effect Thesis', *Philosophical Quarterly*, 19 (1969) pp. 337-42.)

Actions offers not one but three arguments for the view summarized in the sentence Lowe made into the title of his article. Even from a successful criticism of the first of them, he could scarcely reach the strange conclusion, to which he says he is drawn, that actions are not events. I would only add that Chapter Seven says concerning the *ir*reducibility of agent causation that it is 'surely true... and dubious whether it has ever been denied' (p. 96).

Corpus Christi College, Oxford

© JENNIFER HORNSBY 1982

MATERIALISM AND MEMORY

By ANDREW WARD

IN 'Materialism and the Unity of Consciousness' (ANALYSIS 40.3, June 1980, pp. 144-6), I contended that anyone who depends on a certain argument for his right to claim that the material world exists (rather than a mere succession of mental phenomena (perceptions)) cannot *also* justifiably claim that the brain makes possible the accuracy of memory. In 'Memories and the World' (ANALYSIS 41.3, June 1981, pp. 141-5), Mr D. H. M. Brooks takes issue with me. Before responding, I shall set out the two, structurally similar, arguments on the basis of which I put forward my contention.

ARGUMENT I

If there were *not* a material world, it would be highly unreasonable to expect the succession of perceptions to have had the order it has *actually* had, i.e. an order in which successive perceptions are continually so similar as to suggest they are intermittent experiences of an external world of persisting entities. For, if there were a mere succession of perceptions, it would be an inexplicable coincidence that, out of all the logically possible ways a mere succession of perceptions *might* have come, it should *actually* have been so ordered as to suggest that the perceptions *were* successive experiences of an external world of persisting entities.

If there *were* a material world, it would be wholly reasonable to expect the succession of perceptions to have had the degree of stability it has actually had. For the continuous existence of something external to cause the successive perceptions explains why they have been so similar.

Therefore, given the succession *has* actually come in just such a way as to suggest that the perceptions *were* intermittent experiences of continuously existing external entities, it is more reasonable to believe in a material world than a mere succession of perceptions.

(I am not saying that this first argument is valid, only that anyone who does accept it ought also to accept the second argument. I have discussed why I do not think arguments like I refute the sceptic in 'What's not really wrong with Phenomenalism', *Australasian Journal of Philosophy* 1973, pp. 245-52.)

ARGUMENT II

If experiences of given past times were *not* connected with one's present ostensible memory of those past times, it would be highly unreasonable to expect the ostensible memory to be *accurate*, i.e. to *match* those past experiences. For, if there were not a *connection*, it would be an inexplicable coincidence that, out of all the logically possible ways in which a present ostensible memory of given past times *might* have come, the *actual* ostensible memory should *match* the very experiences of those past times.

If the past experiences *were* connected with one's present ostensible memory, it would be wholly reasonable to expect the ostensible memory to match them. For the continuous existence of something to store and reproduce their content explains the matching between the past experiences and the present ostensible memory.

Therefore, given one's present ostensible memory *is* accurate, it is more reasonable to believe that there is, rather than that there is not, a connection between the past experiences and the ostensible memory.

(Brooks seriously mis-states my presentation of II: as he puts it, my contention manifestly could not even get off the ground. For, according to him, II *alone* would apparently prove *both* the existence of the required connection *and* the existence of at least one persisting external entity, the brain. But II cannot be used to prove both because, until one has good reason for thinking that perceptions *have* come in a certain stable order, one has no good reason for claiming that there exists anything external. Argument I attempts to prove the continuous existence of something *external* to the mind; Argument II attempts to prove the continuous existence of some *unifier* of the mind's experiences. But the latter argument leaves *open* the question of whether the unification is or is not made possible by something external to the mind.)

My contention is that anyone who, by depending on I, believes himself justified in claiming that the material world exists is not justified in claiming that anything material makes possible the accuracy of memory. He is not, since the acceptability of I presupposes the accuracy of memory: without this presupposition, one would have no good reason for thinking that the succession of perceptions *has* been caused by a material world. But, by II, unless there is a connection between past experiences and one's present ostensible memory, it would be highly unreasonable to trust the ostensible memory. However, one can have no good reason for thinking the required connection material because memory's accuracy must be presupposed in order to claim that I does prove the existence of matter.

Turning to Brooks' criticisms, he says: (a) I wrongly assume Argument I must use the *actual* order of past perceptions and not the order they *appear* to have through present ostensible memory; (b) I wrongly assume that one of the two arguments must be dealt with *before* the other; in fact, they can be considered simultaneously; (c) When they *are* considered simultaneously, the hypothesis of a material world (including the brain) is seen to be a much more plausible explanation of experiences than any of the suggested alternative explanations.

I take (a) and (c) together. Brooks declares that even if one does not presuppose the accuracy of ostensible memory – and, hence, so far as one is aware, the succession of perceptions *may or may not* have come as ostensibly remembered – one still has some explaining to do: one has to explain the way in which one seems to have had and to be having perceptions. He maintains that the materialist hypothesis is a much better explanation than any other currently available, better, e.g. than the hypothesis that the perceptions have been caused by Descartes' demon or Berkeley's God.

But the point is not 'Given the succession of perceptions *has* an external cause, is the best explanation a material world or, e.g. a malignant demon?'; the point is whether it is more reasonable to accept a mere succession of perceptions or a succession of perceptions *and* an external world of persisting entities to explain the order of that succession. It is a pervasive premise of Brooks' objections that experiences *have* an external cause. But that is undoubtedly an assumption for which the sceptic demands justification. Argument I does not *assume* that whatever the order of perceptions, they must have an external cause; on the contrary, it seeks to *justify* the existence of an external cause in this particular case, i.e. given the *actual* order of perceptions. If one were to assume that, *whatever* their order, perceptions must have an external cause, the fundamental point raised by the sceptic, viz. whether there is justification for believing in an external world, would simply not have been considered. Now the materialist *is* employing I to claim to have answered *scepticism*, not or not primarily some other form of *realism* (e.g. Descartes' demon or Berkeley's God); and I am concerned with whether his use of I has implications for his further claim that the brain makes possible the accuracy of memory. Criticism (c) merely takes for granted the fundamental point raised by the sceptic and to which I tries to provide an answer.

In the light of these remarks, let us examine criticism (a). Let us, therefore, suppose that one does not assume the accuracy of ostensible memory and, hence that, so far as one is aware, the actual succession of perceptions may or may not have been as ostensibly remembered. It is obvious, I submit, that I, suitably modified, can no longer be seen as setting out to *justify* belief in an external cause. For one no longer has any good reason for thinking that the successive perceptions *have* continually been so similar as to suggest that they are intermittent experiences of an external world of persisting entities. It follows, correspondingly, that one no longer has any good reason for claiming that, if there were no external world of persisting entities, it would be an inexplicable *coincidence* that, out of all the possible ways a mere series of perceptions *might* have come, the actual succession should suggest that the perceptions *were* caused by an external world of persisting entities. But once the appeal to a coincidence lapses, one has no good reason for claiming that an external world exists rather than a mere succession of perceptions.

I do not, therefore, see any force in criticism (a): unless the present accuracy of memory is presupposed, one has no good reason for believing that the successive perceptions *have* continually been so similar and, consequently, the appeal to an inexplicable coincidence without an external world of persisting entities – the cornerstone of I – collapses.

According to criticism (b), because I hold that I and II presuppose each other, I assume that if one of them is taken after the other, 'it is vitiated by the fact that it presupposes the other. In fact, in replying to the sceptic, one can consider the arguments jointly'.

I do not hold that the two arguments presuppose *each other*; in particular, I do not hold that II presupposes I. I hold that the acceptability of I presupposes memory's present accuracy, and that anyone who accepts I should *also*, because it is structurally similar, accept II. From which I argue: (i) unless one presupposes a connection between past experiences and present ostensible memory *logically prior* to employing I, it would be highly unreasonable to claim that I proves matter's existence; and (ii) since those who *do* claim I as a proof of matter must accept that it would be highly unreasonable to do so unless the above-mentioned connection is presupposed (because, without that connection, it would be highly unreasonable of them to trust present ostensible memory), it can be shown that they have no right to claim that the brain makes the connection possible.

The argument for (i) is as follows. The connection between past experiences and present ostensible memory must be presupposed *prior* to employing I because if this required connection does not exist, it would (as II shows) be highly unreasonable to trust ostensible memory; and if the trustworthiness of ostensible memory *were* a highly unreasonable assumption, it would be correspondingly highly unreasonable to claim that I proves matter's existence. For I can only be claimed to prove matter's existence assuming the accuracy of ostensible memory, and *that* would be a highly unreasonable assumption if the required connection does not exist.

As to (ii), the argument proceeds as follows. Since I can only be claimed as a proof of matter presupposing the required connection, those who assert that they are justified in thinking something *material* (in fact, the brain) makes possible the required connection cannot thereby claim that I *does* prove matter's existence. For if the brain makes the required connection possible then unless the brain exists, it would be highly unreasonable to trust ostensible memory. It would be highly unreasonable because if the brain makes possible the required connection yet does not exist, the required connection cannot exist either; hence it would be highly unreasonable to trust ostensible memory. Consequently those who maintain *both* that I gives good reason for thinking matter exists *and* that they are justified in thinking the brain makes the required connection possible must presuppose the existence of matter (the brain) in order to claim that I proves matter's existence. For they must accept that unless the brain exists, it would be highly unreasonable to trust ostensible memory and hence highly unreasonable to claim that I *does* prove matter's existence. But if they have to presuppose the existence of matter in order to claim that, by employing I, they have proved the existence of matter, their position is circular. In short, in claiming to be justified in thinking that the brain makes possible the required connection, they are, intentionally or not, claiming to have no good reason for thinking that matter, and hence the brain, exists.

Accordingly, anyone who *does* claim that I gives a good reason for believing in matter cannot justifiably assert that the required connection is made possible by *anything* material.[1]

University of York,
Heslington, York YO1 5DD

© ANDREW WARD 1982

---

[1] I am grateful to the Editor for help in improving the presentation of this paper.

WILLIAMS ON 'THE SELF AND THE FUTURE'

By HAROLD W. NOONAN

IN his paper 'The Self and the Future' (*Philosophical Review*, LXXIX, 1970, now collected in *Problems of the Self*, Cambridge 1973, pp. 46-63) Bernard Williams puts forward an argument in favour of a mentalistic criterion of personal identity, and then, without indicating any way in which he considers this argument to be mistaken, an argument in favour of a bodily continuity criterion of personal identity. The implication of the paper is that there is nothing straightforwardly *wrong* with either argument and hence that a choice between these two types of criterion must of necessity involve a risk – 'that there is room for the notion of a *risk* here', Williams writes, 'is itself a major feature of the problem'.

The argument Williams gives for a bodily continuity criterion of personal identity has recently been criticised by John Perry. I accept Perry's criticisms and would add others (which an interested reader can perhaps glean from my 'Can one thing become two?', *Philosophical Studies* 1978). But Perry says of the argument Williams gives for a mentalistic criterion of personal identity that it 'puts the case for the possibility of bodily transfer about as effectively as it has been put'. (J. Perry, review of *Problems of the Self*, *Journal of Philosophy* 73, 1976). I have the impression that Perry's good opinion of this argument is widely shared; at any rate, I disagree with it. In what follows I explain why.

Williams' argument turns on what two people, whom he calls 'the A-body-person' and 'the B-body-person', are likely to say after a certain sequence of events about the wisdom, assessed from a purely selfish point of view, of choices made by two earlier people, A and B. The A-body-person is someone who has A's body and brain – but A's brain wiped clean of all the information it contained and 'reprogrammed' by means of a fabulous machine with information taken from B's brain. The B-body-person is someone who has B's body and brain but B's brain 'reprogrammed' with information taken from A's brain. The sequence of events after which the A-body-person and the B-body-person make their assessment of the choices of A and B is this process of information transfer between A and B (which Williams refers to as 'the experiment') followed by the torture of either the A-body-person or the B-body-person by the experimenter and the receipt of a cheque for $100,000 by the other. The choices made by A and B concern which of these should get the money and which should get the pain. These choices are made in a purely selfish spirit.

Williams' argument is now that what the A-body-person and the B-body-person are likely to say after the experiment in a series of cases in which A and B make different choices provides strong support for the view that the A-body-person is B and the B-body-person is A, and hence for a mentalistic criterion of personal identity.

In order to see what is wrong with this argument, however, one need only notice that the choices of A and B, made in a purely selfish spirit, are bound to reflect their views on personal identity, however inarticulate these views may be: for each is concerned that *he* gets the money and the other get the pain, and each (Williams is clearly assuming though he does not say) thinks he has reason to believe that his choice will be acted on by the experimenter when he decides whom to reward and whom to punish. But since the A-body-person's brain has been wiped clean and reprogrammed to ensure that the A-body-person is identical with B in all psychological respects, the A-body-person must share B's views on personal identity; similarly the B-body-person must share A's views. Furthermore each can assess the wisdom of the choices of A and B only on the basis of *his own* views on personal identity. Hence each must regard the choice he 'recalls', i.e. the choice made by the person with whom he is psychologically identical, as a wise one, and (whether or not he thinks the choice he 'recalls' is his own choice) he must regard his own choice as a wise one if and only if it is a choice in accordance with the views on personal identity he presently holds, i.e. the views which are reflected in the choice he 'recalls'. Once this is understood the plausibility of Williams' argument evaporates. For insofar as the A-body-person and the B-body-person are likely, in the series of cases Williams imagines, to make statements that appear to provide support for a mentalistic criterion of personal identity, this will only be so if they are themselves believers in such a criterion; any criterion of personal identity, no matter how bizarre, could be given a similar appearance of support by imagining what would be said by people crazy enough to accept it (Williams perhaps shows that he is half aware of this, at least for bizarre *mentalistic* criteria, when he says that if character and 'memory'-impressions appropriate to B were produced in A by a process which ensured that the provenance of the 'memory'-impressions clearly disqualified them from counting as veridical memories of B's life 'I might well be able to elicit from the A-body-person the kind of remarks about his previous expectations of the experiment – remarks appropriate to the original B – which so impressed us in the first version of the story' (*Problems of the Self*, p. 56) – i.e. when the argument for a mentalistic criterion of personal identity was being presented). On the other hand, insofar as the A-body-person and the B-body-person do not accept a mentalistic criterion of personal identity they will not be likely to make statements that even *appear* to support such a criterion.

To confirm this let us look more closely at Williams' argument.

(i) First he supposes that A and B, are believers, however inarticulate, in a mentalistic criterion of personal identity, and choose appropriately: A that the A-body-person get the pain and the B-body-person the money, and B conversely, and that after the experiment the experimenter gives the pleasant treatment to the A-body-person and the unpleasant treatment to the B-body-person. Both the A-body-person and the B-body-person are now bound to regard the choices they 'recall' as wise ones, and each is bound to regard the choice he 'recalls' as his own. Each is therefore bound to believe that his own choice was a wise one and to make exactly the type of remarks Williams imagines. Namely, 'the B-body-person will not only complain of the unpleasant treatment as such but will complain that that was not the outcome he chose' and the A-body-person 'will express satisfaction at the fact that the experimenter has chosen to act in the way he, B, so wisely chose' (*Problems of the Self*, p. 49). But contrary to what Williams claims these facts provide no support at all for a mentalistic criterion of personal identity, for what the A-body-person and the B-body-person are likely to say is not determined merely by their memories, but also by the views they hold on personal identity.

This becomes clearer when we consider the cases where A and B make other choices with regard to the experiment, where Williams is led into inaccuracy in his description of what the A-body-person and the B-body-person are likely to say.

(ii) Next Williams supposes that both A and B make choices appropriate to a belief in a bodily continuity criterion of personal identity and that after the experiment the experimenter gives the pleasant treatment to the A-body-person and the unpleasant treatment to the B-body-person. Each is bound to regard the choice he 'recalls' as a wise one and also to regard his own choice as a wise one.[1] But each is bound to distinguish his own choice from the choice he 'recalls' and to dismiss his 'memory'-impressions as illusory. Hence Williams is wrong when he says that the B-body-person will 'naturally express' his acknowledgement that the distribution of pain and pleasure is in accord with A's choice by saying that this is the distribution he chose. Rather he will merely have a *tendency* to express this acknowledgement in this way (if he had no such tendency he would not have 'memory'-impressions of A's life at all – whether or not illusory), a tendency he will try to suppress: his considered statement will be to the effect that the distribution is not the one he chose, but merely the one he has an illusory 'memory'-impression of choosing. Similarly the A-body-person will not, as Williams says, reckon it good luck that the experimenter did not produce the outcome he recalls (minus scare quotes) choosing, but merely reckon it good luck that the experimenter did not produce the outcome he has an illusory 'memory' of choosing, and rather produced the outcome he chose, but knew before the experiment he would have no recollection of choosing.

---

[1] The A-body-person can turn to the B-body-person and say (and the B-body-person will not dissent) 'You are the B-body-person and chose money for the B-body-person and pain for the A-body-person, and I am the A-body-person and chose money for the A-body-person and pain for the B-body-person; from a purely selfish point of view both your choice, which I 'recall' though you don't, and my choice, which you 'recall' though I don't, were wise ones'.

(iii) Finally Williams supposes that A chooses in a way appropriate to a mentalistic criterion of personal identity and B chooses in a way appropriate to a bodily continuity criterion of personal identity, and that the experimenter produces the outcome they both want, namely money for the B-body-person and pain for the A-body-person. Each of these is now bound to regard the choice he 'recalls' as a wise one, but otherwise they will differ. The B-body-person will regard the choice he 'recalls' as his own choice, i.e. a choice he genuinely recalls, and so regard his own choice as a wise one. The A-body-person, however, is bound to distinguish the choice he 'recalls' from his own choice, and to regard his own choice as a mistaken one.[2] Williams is right, then, when he says that in this case the B-body-person 'likes what he is receiving, recalls choosing it, and congratulates himself on the wisdom of (as he puts it) his choice'; but wrong when he says that the A-body-person 'does not like what he is receiving, recalls choosing it, and is forced to acknowledge that (as he puts it) his choice was unwise'. Of course the A-body-person does not like what he is receiving, but he would not accept that it is what he recalls (minus scare quotes) choosing, nor would he acknowledge that the choice he 'recalls' choosing was unwise, nor would he be prepared to put any comment about the choice he 'recalls' in the form 'My choice...'. Once again, the point that emerges is that Williams accurately describes what the A-body-person or the B-body-person is likely to say only when that person accepts a mentalistic criterion of personal identity – when, however, what he says, being expressive of a commitment to such a criterion, cannot, for just that reason, provide more than an appearance of support for it; while whenever the A-body-person or the B-body-person accepts a bodily continuity criterion of personal identity Williams is forced to misdescribe what he is likely to say in order to preserve any appearance of cogency for his argument.

This becomes even clearer, if anything, when we look at some subsidiary arguments Williams gives. He now drops the supposition that A and B are forced to choose how the A-body-person and the B-body-person are to be treated after the experiment and assumes a more benevolent character for the experimenter. First he supposes that A and B are disposed to accept a mentalistic criterion of personal identity, and that each agrees to the experiment partly, at least, in the hope of exchanging his body for a more attractive one. He then points out, what the reader should now be able to predict, that what the A-body-person and the B-body-person are likely to say after the experiment now appears to provide support for a mentalistic criterion of personal identity. Next he supposes that when the experiment is proposed A and B think rather of their psychological advantages and disadvantages. A's thoughts turn primarily to certain sorts of anxiety to which he is very prone, while B is concerned with the frightful memories he has of past experiences which still distress him. They are each inclined to accept a bodily continuity criterion of personal identity, and so each hopes that the experiment will result in his acquiring a healthier psychological state. A, for example, reasons that if the experiment comes off the person who is bodily continuous with him will not have his anxiety, and while the other person will no doubt have some anxiety – perhaps in some sense his anxiety – at least that person will not be he. The experiment is performed and the experimenter (to whom A and B previously revealed privately their several difficulties and hopes) asks the A-body-person whether he has got rid of his anxiety. Williams now writes 'this person presumably replies that he does not know what the man is talking about; he never had such anxiety, but he did have some very disagreeable memories, and recalls engaging in the experiment to get rid of them, and he is disappointed to discover that he still has them. The B-body-person will react in a similar way to questions about his painful memories, pointing out that he still has his anxiety'. But this, as the reader will now appreciate, is wholly mistaken. Since both A and B accepted a bodily continuity criterion of personal identity, so must the A-body-person and the B-body-person. The A-body-person will therefore reply to the experimenter's question by saying something to the effect that since he does not know what the experimenter is talking about he presumes that he has indeed got rid of the anxiety to which he was so prone, and that from his (A's) point of view the experiment has been a complete success. Unfortunately, he will continue, he has acquired instead some illusory 'memories' of B's life of a distinctly unpleasant character, but, he might add, given B's willingness to engage in the experiment in the first place some such outcome was only to be expected. The B-body-person will react in a similar way to questions about his painful memories, and say he now realizes why A was so willing to participate in the experiment with him (B). Once again then Williams is able to give his argument an appearance of plausibility only by misdescribing what the A-body-person and the B-body-person are likely to say in the situation he puts them into.

---

[2] But his views are likely to change. Since he regards himself as A, he is bound to agree that his views have changed since the time he made his choice; not, however, as a result of a rational change of mind, but rather as a result of his mind's being changed for him. Reflecting on this he seems bound to be led to ask whether he *ought* to regard bodily continuity as necessary for personal identity (particularly if, say, his 'recollection' of A through B's eyes informs him that A thought B a dunderhead with no appreciation at all of the overwhelming arguments in favour of a mentalistic criterion of personal identity), but once he seriously asks this question he ceases to be a believer and becomes an agnostic. The view he eventually comes to accept at his settled opinion may be either A's or B's, so he may eventually come to regard the choice he 'recalls' as his own choice and to accept that that choice was unwise. I hope it is clear that this does not invalidate the criticism of Williams in the immediately following text.

In sum, then, if Perry is right in saying that Williams' argument puts the case for the possibility of bodily transfer (and hence for a mentalistic criterion of personal identify) 'about as effectively as it has been put', supporters of the bodily continuity criterion have been given no good reason at all for modifying their views. Whether Perry is right, however, is another question altogether.

University of Birmingham,
Birmingham BI5 2TT

© HAROLD W. NOONAN 1982

MINCE PIE REASONING

By GERALD W. BARNES

> '... one might easily wonder why no one has ever pointed out the mince pie syllogism...' (G. E. M. Anscombe, *Intention*, 2nd edition 1969, sec. 33)

I THINK it likely that most philosophers would be initially inclined to go along with A. J. P. Kenny's assertion that it 'is beyond doubt that in addition to theoretical reasoning there is practical reasoning'. ('Practical Inference', ANALYSIS 26.3, January 1966, p. 65.) Yet if we understand this assertion in what seems the natural way – viz., as an assertion to the effect that practical reasoning is or involves, at least in part, a sort of reasoning that is not theoretical reasoning – then it turns out, I think, that we ought to take seriously the possibility that the assertion is false. In this paper I try to indicate why we ought to take that possibility seriously.

For someone to do some theoretical reasoning is for a psychological process to occur, a process which embodies some deductive or inductive argument, i.e., some abstract propositional or sentential structure of the relevant sort, and which culminates in a belief. For someone to do some practical reasoning is for a psychological process to occur that culminates in (depending on the account of practical reasoning) a desire, or volition, or intention, or other psychological state with, so to speak, immediate, even if defeasible, motivational efficacy. It is clear enough, I think, that neither of these two sorts of psychological process is simply an example of the other.

Thoughts, feelings, etc. and collections of these exhibit great variety, as do also the ways in which one such can lead to another. It happens sometimes that a psychological process occurs whose upshot is what we could call a grounded belief, and sometimes one whose upshot is what we could call a grounded desire – i.e., the belief that so-and-so on the grounds that such-and-such, or the desire that so-and-so on the grounds that such-and-such. Theoretical reasoning is at least an important variety of psychological process resulting in grounded belief; practical reasoning is an important variety of psychological process resulting in grounded desire (or volition, or intention).

Sometimes one comes to have a grounded desire without going through a process of practical reasoning. Suppose that Ralph has just finished a dinner whose aromas, flavours, etc. have aroused in him a strong desire for some coffee. His host produces a cup of coffee, *c*, that Ralph recognizes as such, and offers it to Ralph. We can expect Ralph at this point to come to have a desire, even if only a prima facie desire, to have *c*, on the grounds that, or because, or for the reason that *c* is a cup of coffee, and also quite possibly to take *c* because of this desire. It seems wrong to call this a case of practical reasoning; no 'calculation' occurs, to borrow Anscombe's term (G. E. M. Anscombe, *Intention*, sec. 35). Yet Ralph does have at least a rudimentary reason for wanting to have *c*, namely, he believes concerning *c* that it is a cup of coffee, in the context of his unreasoned general desire for some coffee.

Regarding cases such as this one, i.e., cases where one comes otherwise than through practical reasoning to have a grounded desire, it seems plausible, I think, to suppose that one's immediate grounds for that desire consist always of an appropriate relatively general desire together with an appropriate relatively particular belief. It is important to note that it does not *follow* that, for example, Ralph wants to have *c* from the fact that be believes that *c* is a cup of coffee together with the fact that he wants to have some coffee. As Davidson has pointed out (in 'Freedom to Act', collected in *Essays on Actions and Events*, Oxford, Clarendon Press 1980, p. 77), there will always have been at least the logical possibility that Ralph might have simply failed to 'put together' the belief and desire that form the grounds on which he wants to have *c*. Ralph could have been inattentive, or distracted, or self-deceived, *vel cetera*.

If there is this logical possibility for any such case, then we have a distinction between two sub-processes within the overall process of one's coming otherwise than through practical reasoning to have a grounded desire. A first sub-process consists of all and only those events within the overall process, whatever they are on a given occasion, that contribute to one's acquiring the two grounding attitudes in question, i.e., acquiring the relevant general desire and particular belief. A second 'putting together' sub-process consists of whatever else it is that happens on that occasion that, together with the first sub-process, contributes to one's coming on that occasion to have the grounded desire, i.e., contributes to the completion of the overall process. The one sub-process is first and the other is second in the sense that the second cannot end before the first does, while the first can end before the second does. The two may or may not overlap in time. If they do not, the second sub-process could have for one who undergoes it the aspect of one's doing a double-take.

Let us use the label 'G-processes' to designate those psychological processes of which the second sub-process in any case of the sort just considered is an example. G-processes can, I think, be as plausibly supposed to occur in connection with one's coming to have grounded attitudes of any other conative sort, as they can in connection with grounded desire. I want now to sketch two hypotheses about G-processes, neither of which, it seems to me, is obviously false. Taken together, these hypotheses have the consequence that Kenny's assertion, as we are understanding it, is false.

Hypothesis I: G-processes are not any sort of reasoning process.

I believe that an elaborate argument could, in fact, be given in support of Hypothesis I. At the heart of this argument would lie the idea that, to put it very roughly, for any case where a G-process occurs, the ground*ed* attitude (grounded desire, grounded volition, grounded intention, or whatever) can be faulted as irrational only if at least one of its ground*ing* attitudes can (independently) be so faulted. This idea, if correct, distinguishes G-processes from reasoning processes, in particular from such reasoning processes as those sub-processes, partly analogous to G-processes, that occur in connection with the overall process of one's coming to have grounded beliefs or other cognitive attitudes. When one reasons to (for example) a grounded belief, that ground*ed* attitude may be irrational even when its immediate ground*ing* attitudes (beliefs) are not irrational – this will be the case when one's reasoning has proceeded in violation of some canon of inference.

It being a lengthy undertaking to fill in the details of this argument, I shall only point out here that the burden of proof would seem to rest upon anyone who would claim that there is some sort of reasoning exemplified by G-processes. In the case of Ralph's coming to want to have *c* on the grounds that *c* is a cup of coffee, for example, the G-process is obviously not a bit of theoretical reasoning on Ralph's part. What reason is there to suppose that the only alternative is that it is a bit of reasoning of some other sort on Ralph's part? There *seem*, certainly, to be non-reasoning processes whereby beliefs and desires give rise to other beliefs and desires – for example, it seems that desiring that *p* can cause one to believe that *p*; or, that one's belief that *p* can cause one to remember that *q*, or to notice that *r*; and so on. If there are further (epistemic) possibilities in addition to the possibility that G-processes are bits of reasoning, then the question of the truth of Hypothesis I is at present quite open.

Hypothesis II: The controversial element of any case of practical reasoning can be understood as a structured collection of processes of two sorts, viz., G-processes and theoretical reasoning.

To begin to clarify Hypothesis II, let us see how it would have us understand a case in which Ralph comes to have a grounded desire for *c* as a result of what looks like practical reasoning. Suppose again that Ralph's host offers him *c*, and that Ralph recognizes *c* as a cup of coffee. There is, let us suppose, already present a general desire on Ralph's part for some coffee, and the relevant G-process occurs, culminating in Ralph's wanting to have *c* on the grounds that *c* is a cup of coffee. This time, however, we introduce a different supposition concerning Ralph's desire for some coffee: rather than an unreasoned desire, stirred up by tastes and smells, it is itself a grounded desire. For instance, it could happen that Ralph wants to be able to drive home safely soon and believes that, owing to his present drowsiness, he will be able to drive home safely soon only if he has some coffee. Hypothesis II suggests that Ralph's grounded desire for some coffee will be a result of a G-process operating upon some such grounding attitudes as the belief and desire just mentioned.

Because Ralph's desire for some coffee is a grounded one, his desire to have *c* begins already to look more like the outcome of practical reasoning than it did in the original case; it looks even more so as the description of the current case is suitably augmented. Ralph's belief that he will be able to drive home safely soon only if he has some coffee can quite naturally be supposed to be the conclusion of a bit of theoretical reasoning on Ralph's part. Hypothesis II suggests that his desire to be able to drive home safely soon could be yet another grounded desire, arising from yet another G-process. For example, it could happen that Ralph wants to get enough sleep to have a clear head for tomorrow's tasks, and believes that he must be able to drive home (safely) soon if he is to have any chance of getting that much sleep; a G-process could occur in which Ralph 'puts together' this belief and desire. Nor need we stop here. The belief just mentioned can be supposed to result from theoretical reasoning, and the desire to be a further grounded desire, underlain by further G-processes and bits of theoretical reasoning.

We have here a description of a purported case of practical reasoning that appeals only to G-processes, theoretical reasoning, and what are for present purposes non-controversial additional events, such as Ralph's recognizing *c* as a cup of coffee. Hypothesis II maintains that this description gives us everything that goes toward making the case a case of practical reasoning. It maintains also that something in all important respects similar can be done for any pre-analytically acknowledged case of practical reasoning.

Hypothesis II and Hypothesis I are, I claim, not obviously mistaken; but unless at least one of them is mistaken, there is no such thing as practical reasoning in Kenny's sense. Let me attempt now to deal with possible objections.

One possible objection is that the picture of practical reasoning drawn by Hypothesis II misses its intended subject: practical reasoning is essentially a matter of reasoning about what it is best to do, or about what one ought to do, whereas evaluations, desirability-characterizations, etc. play no part in the story told by Hypothesis II.[1] The first thing I would say in reply to this is that I doubt that practical reasoning is concerned always with evaluations. We seem to reason sometimes about how to get what we want, regardless of the merits of what we want, and to act on desires, etc. that result in part from this reasoning; I am not persuaded that such cases, taken at face value, should not count as cases of practical reasoning.

Apart from this, however, it does seem undeniable that there exist cases of practical reasoning in which evaluations do play a major role. It seems clear that a case might occur in which it was tempting to say that Ralph's grounded desire for *c* had as its grounds not the belief that *c* was a cup of coffee together with the desire for some coffee, but that belief together with the belief that it would be good to drink some coffee. This last belief might be such as to incline us to view it, too, as grounded in further beliefs, e.g., in a belief as to the desirability of doing something to enable oneself to get home in time for a good night's sleep. There is the appearance, in such a case, of practical reasoning terminating in a process that is neither a G-process nor theoretical reasoning, contrary to Hypothesis II.

I do not concede that the facts, in such a case, correspond to this appearance. Even if they do, however, nothing follows to the effect that the process in question is some kind of reasoning process. The present objection does not, therefore, come to grips with the main issue, even if it succeeds in refuting Hypothesis II. (I would claim, indeed, that an argument much like the argument that shows G-processes not to be reasoning processes would be available to show that no process of the kind now contemplated is a reasoning process, should it turn out that we must concede the existence of processes of this kind.)

A second possible objection might proceed from the intuition(s) that there is such a thing as 'the logic of imperatives', or implication-relations among 'fiats', or something of the sort. For example, it has seemed to some – to Hare, for example, in *The Language of*

---

[1] For recent accounts of practical reasoning in this vein, see, e.g., Davidson's 'How Is Weakness Of The Will Possible?' and 'Intending', *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), pp. 21-42 and 83-102, or M. Bratman's 'Practical Reasoning and Weakness of the Will', in *Nous*, XIII (May, 1979), 153-71.

Morals – that from such 'premises' as 'Shut off all the lights!' and 'That's a light', there follows, in *some* sense of 'follow', the imperative 'Shut that off!'. If this were an argument of some kind, then presumably there would be a variety of reasoning corresponding to this and other arguments involving imperatives. That variety of reasoning would seem to be a significantly different variety, however, from theoretical reasoning, the variety corresponding to purely indicative arguments. It does not seem to hold true in general that an indicative argument form is valid just in case its imperative quasi-counterpart is so. For example, 'p; ∴, $p \vee q$' is valid for indicative sentences, but can seem not to be so for imperatives if we consider 'Shut off the light!; ∴, Either shut off the light or finish your breakfast!'. If there are two sorts of argument, imperative and indicative, that differ so significantly with respect to validity, then it looks as if the two corresponding varieties of reasoning would also have to differ to an extent that would entitle us to speak, with Kenny, of a second variety of reasoning 'in addition to' theoretical reasoning; the two Hypotheses under discussion could not both be true, even if it is not readily apparent exactly where the falsehood lies.

Thus stated, however, this objection overlooks a serious problem concerning the intuitions to which it appeals: they are inconsistent. For example, the form of imperative inference above that Hare accepts is rejected by Kenny (op. cit., p. 74). Perhaps there are intuitions of the relevant sort which a defender of Hypotheses I and II needs to be able to explain away, but it does not yet seem clear which intuitions in particular these are. As long as that remains unclear, this objection is inconclusive.

A third possible objection is that the view implicit in Hypothesis II overlooks a crucial element of practical reasoning, namely, the transition that occurs from the (mere) prima facie desire to do so-and-so to the unconditional desire, or volition, or intention to do so-and-so. This transition is not plausibly viewed as either a G-process or a bit of theoretical reasoning, the objection continues, and so Hypothesis II is false.

The cogency of this objection can be determined only to the extent that we are clear about such matters as how intention differs from prima facie desire, what it is to form an intention, and so on. Here I can only point out that, even if the objection were quite correct, and Hypothesis II were thereby falsified, a defender of Kenny's assertion could take little comfort in the fact without further argument. As noted earlier, it seems that many processes whereby psychological states give rise to other psychological states are not reasoning processes. Even if there is sometimes or always in practical reasoning a transition from a psychological state of one sort, desire, to one of a distinct second sort, e.g., intention, it would remain to be shown that non-theoretical reasoning processes play a part in such transitions.

Trenton State College,
Hillwood Lakes, Trenton,
New Jersey 08625, U.S.A.

© GERALD W. BARNES 1982

INTENTIONALITY: SPONTANEOUS ASCRIPTION AND DEEP INTUITION

By KIM DAVIES

IN 'Killing People Intentionally, By Chance' (ANALYSIS 41.3, June 1981, p. 156), I offered a solution to Butler's problem which consisted of an explanation of why our immediate responses to the two cases differ, together with an account of how both responses can be, but are not always, correct. This solution made use of a difference in the scope of the intentions normally assumed to be present in the two cases. Thus, normally an intention to kill someone is not limited to killing them under the conditions of a game of chance. Assuming the presence of this more general intention, we spontaneously call the killing intentional; and where such an intention is in fact present, this is correct. In the case of throwing a six, such an intention is not normally present; our spontaneous response is guided rather by the fact that in a game of dice it is all important that what number turns up on the throw is not subject to the thrower's control, and hence is not intentional. Where no general intention concerning the throw of a six is present, this response is correct. Thus we have an explanation of our differing spontaneous ascriptions of intentionality, and an account of the conditions under which they are both correct. Eric Stiffler, however, in his 'Butler's Problem Again' (ANALYSIS 41.4, October 1981, p. 216), objects that this solution depends on the claim that throwing a six in an ordinary dice game cannot be instantiated intentionally, and that this claim is false. Brown can have the intention to throw a six in an ordinary dice game *at some time or other* (he may set up dice games until he does so). Here it would be intentional that he throw a six, without it being intentional that *on this particular occasion* he throw a six. This objection misunderstands the nature of the solution. In showing how it can be correct to call the killing of Smith intentional, and correct to call the throw of the six not intentional, we have answered precisely Butler's question (presented in ANALYSIS 37.3, March 1977, p. 97). But this does not imply that the normal conditions on which these ascriptions are founded always obtain. Where they do not — as where Brown lacks a general intention to kill Smith, or possesses a general intention to throw a six in an ordinary dice game – these ascriptions are incorrect.[1]

This will not satisfy E. J. Lowe. In his 'Intentionality and Intuition: a reply to Davies' (ANALYSIS 42.2, March, 1982, p. 85),[2] he claims that Butler must have meant to talk of the intentionality that *on the particular occasion* Brown did something, or else the problem rests on an equivocation and is bogus from the start. This seems a bit hard-hearted. It is not clear that Butler had any precise thoughts at all in this respect. Yet even without such thoughts, there remains the problem of how our spontaneous ascriptions of intentionality can differ in the two cases, and yet both be correct: the problem I claim to have solved. Still, if the problem is cast in Lowe's light, then it is clear that our immediate response to Brown's killing Smith is incorrect. It is not intentional that, on that particular occasion, Brown kill Smith, since, having pulled the trigger, he has no control over whether the bullet is fired. Surely then, argues Lowe, there is no disagreement between us, since he also thinks that this ascription of intentionality is incorrect (see his 'Peacock and Kraemer on Butler's problem', ANALYSIS 40.3, June 1980, p. 113).

There is a disagreement between us. To clarify it, I want to distinguish between spontaneous ascriptions and deep intuitions. To grasp the distinction, let us see it at work elsewhere in philosophy. Consider the case of parents who secretly kill their new-born child. Let us assume that the child would have had a life in which, on utilitarian calculation, good would exactly have balanced harm, and that the parents will be far happier without the child than with it. Our immediate response to this may well be that it is morally impermissible to kill the child; a response based on the intuitively held principle that it is wrong to kill an innocent person. Now consider two possible rejections of this spontaneous ascription of moral impermissibility. The first, on the part of a certain kind of utilitarian, rejects *also* the intuitively held principle, and argues that since utility will be maximized by killing the child, it is at least permissible to do so. The second argues that the child is not a person in the morally relevant sense since, say, it does not possess the attribute of self-awareness. Here the spontaneous ascription is rejected, while the deep intuition concerning the wrongness of killing innocent people is retained. It is just that this situation departs from the normal conditions of application of that intuition, the conditions which explain the spontaneous ascription. In this more complex case, the spontaneous ascription is held not to be justified. It is clear that agreement regarding this particular spontaneous ascription is consistent with disagreement on the deep intuition on which that ascription is based, and with disagreement on other ascriptions of moral permissibility in other cases.

---

[1] With regard to Stiffler's view that I argue, in effect, that throwing a six in an ordinary dice game cannot be instantiated intentionally, I should mention that in an earlier draft of my paper, shortened on the Editor's advice, I had discussed precisely Stiffler's example. It was not, then, my intention that the argument have this effect.

[2] I am grateful to Jonathan Lowe for letting me see this paper prior to publication.

The present situation is analogous. Lowe and I agree in rejecting the spontaneous ascription of intentionality in the case at issue (understood on Lowe's terms). I reject it whilst retaining what I take to be our deep intuition concerning intentionality, namely that doing something is intentional if it is done because of an intention on the agent's part to do it. It is just that in this unusual case, the conditions in which the intuition is normally applied do not obtain, so that the spontaneous ascription based on this intuition is not here justified. Lowe rejects the spontaneous ascription, but also the deep intuition. He holds an account of intentionality which makes no reference to the agent's intentions.[3] This account leads him to disagree with ascriptions of intentionality based on the deep intuition; for example, even with a general intention to kill Smith, Brown's killing is not intentional, but rather 'neither intentional nor unintentional'.[4] This is not the place to discuss Lowe's general analysis of intentionality. It is enough to have shown that our spontaneous ascriptions of intentionality in Butler's problem can coherently be justified. Even where these ascriptions are not tenable, no doubt is cast on the soundness of that deep intuition, whose reflective application itself justifies their revision.

Sunderland Polytechnic, Chester Road,
Sunderland SR1 3SD

© KIM DAVIES 1982

---

[3] See E. J. Lowe, 'An Analysis of Intentionality', *The Philosophical Quarterly*, vol. 31, October 1981.

[4] Since writing this, Jonathan Lowe has told me of his uncertainty as to the application of his analysis to cases which involve these general intentions.

AGAINST A DEONTIC ARGUMENT FOR GOD'S EXISTENCE

By PATRICK GRIM

CARL R. KORDIG has recently presented the following two-part argument for the existence of God [4]:

I Og
$Og \supset \Diamond g$
$\therefore \Diamond g$
II $\Diamond g$
$\Diamond g \supset \Box g$
$\therefore \Box g$

Here 'g' is to stand for 'God exists' and 'God' is to be defined as 'a most perfect being'.

Part II of Kordig's argument is a form of the ontological argument similar to that presented by Hartshorne [3] and Plantinga [5], [6]. But part I attempts to supply what many have thought has been crucially missing from ontological arguments – a supplementary argument for the *possibility* of God's existence. (See for example [1], [2], [8] and [9].) Kordig attempts to argue for the possibility of God's existence, moreover, on deontic grounds alone.

Kordig's deontic argument, however, suffers from the same general defect as its ontological predecessors: it fails to get off the ground.

Consider the first premise: 'Og', or 'It ought to be the case that God exists'. Kordig maintains that even atheists 'should grant that a perfect being *ought* to exist' (p. 207). But why? Surely one simple atheistic strategy would be to insist that it is not possible that God exists – just as the atheist has presumably maintained in the face of simpler ontological arguments – and to deny 'Og' on the grounds that it is not the case that the impossible ought to exist. One man's *modus ponens* is another man's *modus tollens*.

Let us put aside this standard objection, however, in order to concentrate on some major difficulties in the supplementary argument Kordig offers for his first premise:

> Define God as a most perfect being. What is most perfect is, in particular, deontically perfect.... What is deontically perfect, ought to exist. Thus, God ought to exist. ([4]: p. 207).

How, in particular, are we to read the third sentence of this argument: 'What is deontically perfect, ought to exist'?

One plausible reading is this: that whatever is specified (or at least consistently specified) as deontically perfect ought to exist. But this would be an unacceptable principle. It would, to begin with, commit us to saying that each of a bizarre range of beings ought to exist: deontically perfect and polka dot beings, deontically perfect beings avidly devoted to stamp collecting, and so on. Such a principle would also lead to contradiction. For consider two purported beings, A and B. A is specified as the only deontically perfect being, as necessarily existent, and as blue. B is specified as the only deontically perfect being, as necessarily existent, and as non-blue. On the principle at issue, each of these ought to exist, since each is specified as deontically perfect. Since $Op \supset \Diamond p$ for any p, the existence of each is possible. But neither is contingent, so both *must* exist. Given our full specifications for A and B, however, both cannot exist.

If this is the reading intended in Kordig's argument for his first premise, then, that argument offers no reason for the atheist to embrace 'Og'. On this reading that argument relies on a principle which leads to contradiction, and so deserves to be rejected by theist and atheist alike.

We might also offer a simpler reading for this crucial step in Kordig's argument for his first premise, however, which does not lead to contradiction: that it ought to be the case that a deontically perfect being exists, in the simple sense that there ought to be such a being.

There *is* a truth of deontic logic which is close to this, and which even the atheist cannot deny: that a deontically perfect *situation* or a deontically perfect *world* ought to exist. (There are serious difficulties with this notion, however: see [7].) *Perhaps* it can also be shown that a deontically perfect world must contain a deontically perfect being, and the atheist thus be forced to concede that there ought to be a deontically perfect being.

Would this be enough for Kordig's argument? No. That the argument requires more than merely deontic perfection is clear from the second step of the second part of the argument: $\Diamond g \supset \Box g$. Here Kordig says:

> A most perfect being, however, is not contingent: $\sim(\Diamond g \,\&\, \Diamond\sim g)$. For necessary existence is more perfect than contingent existence. Therefore, if it is logically possible that God exists, it is logically necessary that God exists: $\Diamond g \supset \Box g$. ([4]: p. 208)

The logic of the argument does not strictly demand that the being at issue be defined as 'most perfect', however. What *is* required is that He be deontically perfect – so as to lend plausibility to the first premise – and necessary, for the sake of '$\Diamond g \supset \Box g$'.

These two requirements, however – of deontic perfection and of alethic necessity of existence – appear to be logically distinct. Once the being at issue is specified not only as deontically perfect, then, but as necessary as well, the first premise becomes less than compelling and Kordig's argument for it becomes sadly incomplete.

The atheist, we will suppose, must concede that there ought to be a deontically perfect being. Must he also concede that there ought to be a deontically perfect *and necessary* being? Kordig offers no reason to think so, and the answer would appear to be 'no'. The atheist is no more forced to say that there ought to be a deontically perfect and necessary being, it appears, than he is forced to say that there ought to be a deontically perfect *and polka-dot* being, or a deontically perfect being who is also an avid stamp collector. Deontic perfection, after all, is perfection in terms of what is permissible and obligatory, and *moral* perfection is its prime exemplar. Existence in many possible worlds, or in all, does not itself seem to be anything like a moral virtue, and it does not appear that any being must fall short of moral perfection or anything like it – must fail to fulfil any duties, for instance, or must violate any rights or commit any wrongs – merely because that being is contingent.

The atheist's commitment to the claim that it ought to be the case that a deontically perfect being exists, then, does not appear to force him to accept 'Og' in any sense that entails that a deontically perfect *and necessary* being exists, and thus does not appear to force him to accept 'Og' in any sense full enough for the purposes of Kordig's argument as a whole.

On one interpretation of Kordig's argument for his first premise, that argument relies on a principle which leads to contradiction. On an alternative interpretation which avoids contradiction, the argument is sadly incomplete. I see no plausible reading of the argument, and no plausible argument to the same purpose, which avoids both of these flaws.

Kordig maintains that even atheists 'should grant a most perfect being *ought* to exist' (p. 207). Whatever its edificatory value, it does not appear that this 'should' has any important *logical* force. The atheist seems no more compelled to grant 'Og' than '◇g', and thus Kordig's deontic argument seems to offer no appreciable advance over its ontological predecessors.

State University of New York,
at Stony Brook, NY 11794, U.S.A.

© PATRICK GRIM 1982

REFERENCES

[1] Patrick Grim, 'In Behalf of "In Behalf of the Fool"', forthcoming in *International Journal for Philosophy of Religion.*
[2] ——, 'Plantinga's God and Other Monstrosities', *Religious Studies*, 15 (1979), 91-97.
[3] Charles Hartshorne, *The Logic of Perfection* (La Salle, Illinois, 1962).
[4] Carl R. Kordig, 'A Deontic Argument for God's Existence', *Nous*, 15 (1981), 207-208.
[5] Alvin Plantinga, *God, Freedom, and Evil* (New York: Harper and Row, 1974).
[6] ——, *The Nature Of Necessity* (Oxford: Clarendon Press, 1974).
[7] Richard Purtill, 'Deontically Perfect Worlds and *Prima Facie* Obligations', *Philosophia*, 3 (1973), 429-438.
[8] ——, 'Plantinga, Necessity, and God', *New Scholasticism*, 50 (1976), 46-60.
[9] Michael Tooley, 'Plantinga's Defense of the Ontological Argument', *Mind*, 90 (1981), 422-427.

ON BEING A CAT

By E. J. LOWE

IN 'The paradox of the 1,001 cats' (ANALYSIS 42.1, January 1982), I argued that a lump of feline tissue *c* could not *be a cat*, on the grounds that the sortal terms 'lump of feline tissue' and 'cat' have different criteria of identity associated with them. To this Professor P. T. Geach replied, in effect, that something may *be a cat* without necessarily complying with the criterion of identity associated with the sortal term 'cat' (ANALYSIS 42.1, January 1982). My objection to this, however, is that it can only be accepted if one also accepts (as Geach does not) that the predicate '—— is a cat' is *ambiguous.*

My reasoning is simple. Contrast the sentences 'Tibbles is a cat' and '*c* is a cat' (where '*c*' denotes a certain lump of feline tissue). Let us grant that both sentences *are* true in *some* sense. It none the less must be the case that they are true in *different* senses. For the implication of saying that Tibbles is a cat is clearly that Tibbles complies with the criterion of identity associated with the sortal term 'cat' (and thus, for example, that Tibbles will not cease to exist merely upon the removal of a single hair). But, evidently, no such implication can attach to saying that *c* is a cat, in any sense in which this can be interpreted as true (since, for example, *c will* cease to exist upon the removal of a single hair). So it seems that we have here *two* different senses of '—— is a cat', one of which demands that the subject of this predicate names something which complies with the criterion of identity associated with the sortal term 'cat', and the other of which does not. But once this is granted Geach's theory becomes untenable, and we are driven towards one like David Wiggins's, which distinguishes between an 'is' of identity and an 'is' of constitution (see his *Sameness and Substance*, Blackwell 1980, p. 30): *c* is a cat only in the sense that *c constitutes* a certain cat, viz., Tibbles. (Leslie Stevenson offers a somewhat different argument for the same conclusion in his 'The Absoluteness of Identity', *Philosophical Books*, XXIII no. 1, January 1982 (a discussion of Wiggins's book and of Harold Noonan's *Objects and Identity*, Nijhoff 1980).)

Is there any way in which a defender of Geach could respond to this objection? What seems undeniable is that 'Tibbles is a cat' implies that Tibbles complies with the criterion of identity associated with the sortal term 'cat', while '*c* is a cat', in any sense in which it can be interpreted as true, does *not* imply that *c* complies with this criterion. Now, I have assumed that the difference between the implications of these two sentences must be attributed to an ambiguity in the predicate '—— is a cat'. But could it not instead be attributed to a difference in sense between the two names 'Tibbles' and '*c*'? (This suggestion was put to me in discussion by Stephen Read.) After all, it may be said, any proper name has associated with it the criterion of identity associated with a certain sortal term, the implication being that the bearer of that name must comply with that criterion. So it may be pointed out that 'Tibbles' and '*c*' differ precisely in that the former has associated with it the criterion of identity associated with the sortal term 'cat', while the latter has associated with it the criterion of identity associated with the sortal term 'lump of feline tissue'. Now, may it not be argued that the reason why, for instance, 'Tibbles is a cat' implies that 'Tibbles' names something complying with the criterion of identity associated with the sortal term 'cat' has nothing to do with the sense of the predicate '—— is a cat', but everything to do with the sense of the proper name 'Tibbles' (and the same, *mutatis mutandis*, as regards '*c* is a cat')? If so, it would seem to follow that '—— is a cat' as it appears in these two sentences is *not* ambiguous.

My answer to this line of argument is as follows. I *accept* that the proper names 'Tibbles' and '*c*' have associated with them the criteria of identity associated with the sortal terms 'cat' and 'lump of feline tissue' respectively. This means that '*c*', for instance, can only be taken to name something which complies with the criterion of identity associated with the sortal term 'lump of feline tissue'. Thus '*c*' cannot, in particular, be taken to name something which complies with the criterion of identity associated with the sortal term 'cat'. *But what does this last assertion amount to*, if not to saying that (in *some* sense) *c is not, and cannot be, a cat*? Surely this is a perfectly legitimate and intelligible way of putting the point; but if so, the clear implication is that we have here a distinctive use of the predicate '—— is a cat', which must be different from any use it can have in saying truly that *c is a cat*. For we may take it as a consequence of a surely indisputable general logical principle that if the sentences '*c* is a cat' and '*c* is not a cat' may both be taken as true, they may, on pain of contradiction, only be so taken if the predicate '—— is a cat' may be used in two different senses.

What I am arguing, thus, is that the very fact that proper names *do* have associated with them the criteria of identity associated with certain sortal terms, far from subserving the Geachian view that the predicate '—— is a cat' is univocal, actually undermines that view. The point is that it is not as though (most) proper names wear their associated criteria of identity on their sleeves. ('Tibbles' is perhaps an atypical example in this respect, in that it is, purely by convention, rarely used to name *anything but* a cat.) Therefore we *need* a way of conveying to others *which particular sortal term's* criterion of identity is associated with the use of a given proper name, say '*N*'. Suppose the sortal term in question is 'cat': then what *better* way of conveying this information is there than by saying '*N* is a cat'? Natural language must possess a relatively simple way of conveying this important information, and there is, I think no good reason to doubt that in English it is done by a certain use of the predicate '—— is a cat'. But if so, and if it is *also* claimed that a sentence like '*c* is a cat' may be interpreted as true even though '*c*' does *not* have associated with it the criterion of identity associated with the sortal term 'cat', but another and incompatible criterion, it must inevitably follow that the predicate '—— is a cat' is *not* being used in this latter sentence in the same way as it was in the sentence '*N* is a cat', and hence that the predicate is, contrary to the Geachian view, ambiguous. For, clearly, in the sense of the predicate '—— is a cat' in which it is used to convey the information that a proper name figuring as its subject has associated with it the criterion of identity associated with the sortal term 'cat' (i.e., as it was used in our sentence '*N* is a cat'), it must be true to say that *c* is *not* a cat, given that '*c*' does *not* have that criterion associated with it. And then it follows, in consequence of our general logical principle, that '—— is a cat' must have two different senses. (My own view, incidentally, is that the predicate '—— is a cat', as it is used in a sentence like 'Tibbles is a cat', is used to say what *sort* or *kind* of thing an individual is, and I call such a sentence an *instantiation* sentence: see, e.g., my 'Laws, Dispositions and Sortal Logic', *American Philosophical Quarterly* 19, No. 1 (January 1982).)

University of Durham,
50 Old Elvet, Durham DH1 3HN

© E. J. LOWE 1982

A THESIS REFUTABLE BY A SENTENCE VERIFIABLE BY ITS USE

By JENNIFER HORNSBY

CONSIDER the following exchange:

> *Goldman:* No action is a person's exemplifying of more than one act property.
> *X:* But my delivering this sentence is my exemplifying at least two act properties.
> *Lemmon:* X's sentence is verifiable by its use.

If Lemmon is right, then what X says is true. But X has contradicted Goldman's statement of his thesis about the individuation of actions.[1] So if Lemmon is right, X has refuted Goldman's thesis.

Is Lemmon right? This was his definition:

> A sentence S is verifiable by its use iff there are circumstances and a manner of delivery such that it is analytic that, for all people *x*, if in those circumstances *x* delivers S in the given manner, then what *x* delivers is true.[2]

The definition confirms that X's sentence is verifiable by its use. The circumstances which guarantee that if X delivers his sentence in an assertoric manner then what he delivers is true are actual circumstances. For anyone uttering X's sentence assertorically actually contradicts Goldman. If so, an act property that X exemplifies besides the property of *delivering his sentence* is the property of *contradicting Goldman*. X was exemplifying at least two act properties, just as he said.

---

[1] See the first two Chapters of Goldman's *A Theory of Human Action* (Princeton University Press 1970). I have used 'action' where Goldman would use 'act token'. Act tokens, Goldman thinks, can be identified *via* triples of persons, times, and act properties.

[2] E. J. Lemmon, 'On Sentences Verifiable by Their Use', *Analysis* 22.4, March 1962, pp. 86-9.

'How can one refute a philosopher simply by denying what he says?' Well, the underlying assumption is that X's sentence is a well-formed bit of English. Any doubts about this will be doubts about Goldman's terminology. But I think that there need not be doubts about this because I think that we can understand the notion of a person's exemplifying an act property on the basis of our understanding of a person's doing something. We could imagine X speaking less technically, then, using the sentence 'My saying what I am is my doing at least two things.'[3] But if (as I believe) Goldman has an erroneous view of what it is to do something, which is at odds with the ordinary use of English words, then, granted that one can use words to do things, it is not surprising if there are speech actions which demonstrate Goldman's error. I claim that X's is one such speech action. Performatives provide other examples.

Corpus Christi College, Oxford ©JENNIFER HORNSBY 1982

---

[3] I introduce 'do something' to give an informal account of what I take to be wrong with Goldman's thesis. Of course one has to keep to his terminology to apply the argument *ad hominem*.

Philosophy

Published for the Royal Institute of Philosophy

Edited by RENFORD BAMBROUGH
President of St. John's College, Cambridge

The journal is concerned with the study of philosophy in all its branches: logic, metaphysics, epistemology, ethics, aesthetics, social and political philosophy and the philosophies of religion, science, history, language, mind and education.

The journal is not committed to any particular school or method and contributors are expected to avoid needless technicality. There is a section on new books which includes reviews, book notes and a list of books received.

Volume 57: January, April, July and October 1982
£40.00/$110.00 per year Single parts £7.50/$20.50

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU — 32 East 57th Street, New York, NY 10022

MIND

A QUARTERLY REVIEW
OF PHILOSOPHY
edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

Vol XCI No 383 July 1982

Articles
- **Description and Identification:** B. HARRISON
- **The Third Man's Contribution to Plato's Paradigmatism:** S. WATERLOW
- **Happiness:** J. KEKES
- **The Complete Reality of Substance:** K. RANKIN
- **Responsibility Especially for Beliefs:** M. STOCKER

Discussions
- **Natural Kinds and Human Artifacts:** D. PUTMAN
- **Why Must Homunculi Be So Stupid?:** E. SOBER
- **On the Function of Consciousness:** P. MOTT
- **Is Wittgenstein's Goethe Stock's Goethe?:** J. WESTPHAL
- **Defeasibility and Memory Knowledge:** A. NAYLOR
- **Eliminative Materialism – a Reply to Everitt:** P. SMITH

Critical Notice:
- **G. Baker and P. Hacker, *Wittgenstein: Understanding and Meaning: An Analytical Commentary on the Philosophical Investigations:*** C. RADFORD

Book Reviews

Books Received

Orders to:
Journals Department,
Basil Blackwell Publisher Ltd.,
108 Cowley Road,
Oxford OX4 1JF, England

Annual subscription:
Institutions: £10.00 or $22.50 (Canada $26.00)
Individuals: £8.00 or $18.00 (Canada $21.00)
Single issues: £3.00 or $6.75 (Canada $7.75)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

PUBLISHER. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. US Mailing Agent, Expediters of the Printed Word Ltd, 527 Madison Avenue, Suite 1217, New York, NY 10022, USA. Second Class postage paid at New York.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £11.00 (inland), £13.50 (overseas), $33.00 (USA), $39.50 (Canada); for individuals £7.00 (inland), £8.60 (overseas), $21.00 (USA), $25.00 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year—in January, March, June and October. Orders should be sent to the publisher.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by Whitstable Litho Ltd, Whitstable.

27 MAY 1983

l. 42 No. 4 (New Series No. 195) October 1982

ANALYSIS

Edited by
Christopher Kirwan

CONTENTS

On a challenge by Anderson and Belnap — Frank Jackson *and* Lloyd Humberstone

Negative existential propositions — Barry Miller

Reply to Miller — C. J. F. Williams

Logical equivalents and logical form — A. J. Dale

Token resistance — Peter M. Simons

Intuitionism disproved? — T. Williamson

Ruben on Lewis and causal sufficiency — Stig Alstrup Rasmussen

Knowing, believing, and guessing — Roy A. Sorensen

Going nowhere fast? — Robert Elliot

Actions, reasons and causal explanation — Denise Meyerson Taylor

Bad news for anomalous monism? — Peter Smith

Reply to Forbes — William Seager *and* Graeme Hunter

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003–2638

Basil Blackwell

Metaphysics

Methods and Problems

GEORGE N. SCHLESINGER

What is metaphysics? What issues and techniques of analysis distinguish metaphysics from the empirical sciences? Where does analysis end and belief begin? In this general introduction to the scope and nature of metaphysics, Professor Schlesinger answers these questions in terms comprehensible by undergraduates.
(November) 288 pages, hardback c. **£17.50** (0 631 13124 8)
paperback c. **£6.95** (0 631 13125 6)

Wittgenstein

G.H. von WRIGHT

Professor von Wright, a pupil and friend of Wittgenstein, has collected together eight studies of Wittgenstein's life and writings. They record the reliable facts of his life and retrieve all the information which is still available relating to the chronology and composition of his writings. Von Wright avoids interpretation and concentrates on providing source material which will be of permanent value to anyone studying Wittgenstein.
224 pages, c. **£12.00** (0 631 13099 3)

Last Writings

on the Philosophy of Psychology

Volume I: Preliminary Studies for Part II of the Philosophical Investigations

LUDWIG WITTGENSTEIN

Edited by G.H. von WRIGHT and HEIKKI NYMAN

This first volume was written between October 1948 and March 1949, when Wittgenstein was having one of his most fruitful working periods, and constitutes the source material for Part II of the Philosophical Investigations.
288 pages, **£19.50** (0 631 12895 6)

Berkeley

New Edition

G.J. WARNOCK

'His exposition and criticism are exact, coherent and lucid.' Times Literary Supplement

'It could hardly be bettered. The book is written with clarity, conciseness, humour and precision.' Spectator

Berkeley is one of the most influential and yet most misunderstood of eighteenth-century philosophers. In this new, revised edition of his classic introduction, G.J. Warnock examines all Berkeley's major philosophical works and discusses his most original and interesting contributions to questions still debated by philosophers today.
240 pages, hardback **£15.00** (0 631 13097 7)
paperback **£5.95** (0 631 13119 1)

Basil Blackwell Publisher, 108 Cowley Road, Oxford OX4 1JF

Basil Blackwell

Wittgenstein on Rules and Private Language

SAUL A. KRIPKE

In this trenchant inquiry Saul Kripke brings his powerful philosophical mind to bear on Wittgenstein's analysis of the notion of following a rule, believing that it is central to his later philosophy. Kripke holds that Wittgenstein's celebrated 'private language' argument, whose interpretation and validity have been the subject of much philosophical debate, is best interpreted in terms of this question.
144 pages, c. **£9.50** (0 631 13077 2)

The Shaping of Man

Philosophical Aspects of Sociobiology

ROGER TRIGG

In a wide-ranging survey of this controversial subject Roger Trigg examines the answers offered by sociobiology, and discusses the contributions of scientists, philosophers, theologians and psychologists as they attempt to explain human behaviour. He argues that human beings are more than the product of their genes and their societies.
206 pages, hardback **£12.50** (0 631 13023 3)
paperback **£5.95** (0 631 13028 4)

Dialogues on Morality and Religion

Selected Translations from 'Julius und Evagoras'

JAKOB FRIEDRICH FRIES

Edited by D.Z. PHILLIPS

These dialogues are the central parts of what Fries called a 'philosophical novel'. In them he wanted to show that the fundamental issues of philosophy – questions, for example, about the point of human existence – could be discussed in ordinary language.
Values and Philosophical Inquiry
272 pages, c. **£18.00** (0 631 10071 7)

Women, Reason and Nature

Some Philosophical Problems with Feminism

CAROL McMILLAN

Accepting that there are irreducible biological differences between women and men, Carol McMillan sets out to disprove the philosophical assumption that these differences entail inferiority or superiority. The book shows that all human activity presupposes reasoned behaviour, and is a forthright assault on all the popular misconceptions which ultimately lead to an undervaluing of women.
Values and Philosophical Inquiry
176 pages, **£12.50** (0 631 12496 9)

Basil Blackwell Publisher, 108 Cowley Road, Oxford OX4 1JF

Basil Blackwell

Plato: Hippias Major

Translated with Commentary and Essay by
PAUL WOODRUFF

Now accepted as genuine and as a crucial text in Plato's development of the theory of forms, Hippias Major here appears in a new translation by Paul Woodruff, accompanied by an extended essay discussing the dialogue as a philosophical whole, and examining its main themes.
(December) 224 pages, c. **£10.00** (0 631 13091 8)

Modern French Marxism

MICHAEL KELLY

This is the first full-length study of French Marxist philosophy and social theory, from its origins to the present day. It provides a critical assessment of the way in which French Marxists have tried to understand human thought in relation to the world, and shows the historical circumstances which have shaped their efforts.
248 pages, c. **£15.00** (0 631 13202 3)

The Ethics of Environmental Concern

ROBIN ATTFIELD

The Ethics of Environmental Concern sets out to examine traditional attitudes to nature and to discover how far these need to be revised to cope with modern ecological problems. It is a scrupulous survey of moral traditions that are often misrepresented and a lively engagement with urgent contemporary issues.
(November) 228 pages, c. **£15.00** (0 631 13137 X)

Rationality and Relativism

Edited by MARTIN HOLLIS and STEVEN LUKES

The book takes its inspiration from the debate encapsulated in **Rationality** edited by Bryan Wilson (1970), but extends it into new fields and into current discussion. Martin Hollis, Steven Lukes and Ernest Gellner contribute the developments in their own thought since the seminal articles included in that book, and are joined by Ian Hacking, W. Newton-Smith, Charles Taylor, Jon Elster, Dan Sperber, Barry Barnes and David Bloor.
320 pages, hardback c. **£16.00** (0 631 12773 9)
paperback c. **£6.50** (0 631 13126 4)

Morality and Language

G.J. WARNOCK

These essays on a number of important topics related to epistemology, morality and language are models of philosophical reasoning and lucid exposition. Among the themes examined are truth, causation, liberty and ordinary language.
(November) 240 pages, c.**£17.50** (0 631 13098 5)

Basil Blackwell Publisher, 108 Cowley Road, Oxford OX4 1JF

Basil Blackwell

New in paperback

On Thinking

GILBERT RYLE

'Like all Ryle's writings, clear, entertaining and instructive.' Gilbert Harman, Times Literary Supplement
152 pages, paperback **£5.25** (0 631 13044 6)

Education and Inquiry

JOHN ANDERSON
Edited by D.Z. Phillips

'The philosophy and practice of education . . . require what Anderson sought to give them – a conception of culture, of criticism, of the nature of progress in knowledge and capacity for argument . . .' From the introductory essay by Eugene Kamenka
240 pages, paperback **£5.25** (0 631 13051 9)

Marxism and Individualism

D.F.B. TUCKER

'This lively, scholarly and contentious book will provoke more thought on what is already a subject of great interest to scholars and students . . . It ranges over a wide spectrum of political, ethical and legal theory . . .' C.B. Macpherson
266 pages, paperback **£5.25** (0 631 13056 X)

Wittgenstein's Lectures, Cambridge 1930 – 1932

From the notes of John King and Desmond Lee
Edited by DESMOND LEE
144 pages, paperback c. **£4.50** (0 631 13167 1)

Wittgenstein's Lectures, Cambridge 1932 – 1935

From the notes of Alice Ambrose and Margaret Macdonald
Edited by ALICE AMBROSE
236 pages, paperback c. **£4.95** (0 631 13166 3)
'It would be difficult to exaggerate the value of these lecture notes for the student of Wittgenstein. They are rich in imaginative detail, often dramatic, and have the ring of authentic genius at work . . . important contributions to our understanding of Wittgenstein's thought and thus make important contributions to learning.' Philosophical Investigations

---

Basil Blackwell Publisher, 108 Cowley Road, Oxford OX4 1JF

ANALYSIS 42.4 OCTOBER 1982

ON A CHALLENGE BY ANDERSON AND BELNAP

By FRANK JACKSON *and* LLOYD HUMBERSTONE

SUPPOSE someone insists that an argument '$P_1, P_2, \therefore C$' is valid, but it turns out that whenever they actually hold P_1 and are faced with the question of inferring C on learning P_2, they refuse to do so *unless* they also hold P_3. This would invite the thought that we are dealing here with an enthymeme and in fact it is '$P_1, P_2, P_3, \therefore C$' which is valid. Alan R. Anderson and Nuel D. Belnap have made a point of this kind against *Disjunctive Syllogism*. Our contention will be that despite its initial appeal it fails, but first we need a little background.

They hold that *Extensional Disjunctive Syllogism* (*EDS*), i.e. '$P \vee Q, {\sim}P, \therefore Q$' is invalid, while granting that *Intensional Disjunctive Syllogism* (*IDS*), i.e. '$P \underline{\vee} Q, {\sim}P, \therefore Q$', is valid, where '$\underline{\vee}$' stands for some intensional 'or'. Their mark of the intensional 'or' is the supporting of subjunctive conditionals with antecedents involving the negation of one disjunct and consequents involving the other disjunct, un-negated. This is deliberately vague, for they do not claim to have yet arrived at a precise criterion for the intensional 'or'.

Now consider the following passage from their *Entailment*, vol. 1, Princeton, 1975, p. 177, where they make a point of the kind in question.

> But whether or not the reader is in sympathy with our [anti-*EDS*] views, it might still be of interest to find a case (if such exists) where a person, other than a logician making jokes, *seriously* holds a proposition *A-or-B*, in a sense warranting inference of *B* with the additional premiss not-*A*, but is unwilling to admit *any* subjunctive conclusion from *A-or-B*. If no such examples exist, then we will feel we have made our case...

We accept that 'no such examples exist'. More particularly, we accept that every case where someone holds *A-or-B* and is warranted in inferring *B* on being given not-*A* is a case where that person holds some subjunctive conditional containing not-*A* in its antecedent and *B* in its consequent. Anderson and Belnap thus have here a *prima facie* case against *EDS* of the kind adverted to in the first paragraph.

Nevertheless it fails. For the subjunctive phenomenon they rightly allege *is precisely what we should expect if EDS is valid*. We will show this by assuming that *EDS* is valid and showing that this assumption, together with doctrines that are widely accepted by classical and relevance logicians alike, *predict* the subjunctive phenomenon in question.

Our argument is not a *petitio principii*. We are not illegitimately assuming *EDS*. The point at issue is the significance of a certain phenomenon involving subjunctives which Anderson and Belnap correctly point out is associated with the employment in practice of Disjunctive Syllogism. If we can show that this phenomenon is predicted, and so can be explained, on the assumption that *EDS* is valid, we will have shown that it is no evidence that *EDS* is invalid. (Had Uranus' perturbations been predictable without Neptune, they would have been no evidence for Neptune.) We will not, of course, have shown that *EDS* is valid, only that the phenomenon in question is no evidence that it is not.

We now give our argument. Anderson and Belnap are considering cases where a person (seriously) *holds A-or-B* and takes it that he or she is warranted in inferring *B* with the additional premise not-*A*. This takes us beyond the question simply of validity. For everyone admits, whatever their view about how to analyse validity, that the fact that '$X, \therefore Y$' is valid is not enough to warrant inferring *Y*. In order for a person to be warranted in inferring *Y*, the argument must be at least *probabilistically sound*; that is, in addition to being valid its premise must be highly probable for the person in question (or, as Anderson and Belnap put it, the person *holds* the premise). In the case of multi-premised arguments, the condition is that the conjunction of *all* the premises be highly probable (which, of course, is different from each premise being individually highly probable). This point has important consequences in the present context. Suppose (i) '$X, Y, \therefore Z$' is valid, and (ii) *X* is highly probable for you (you 'seriously hold' *X*). This is, of course, *not* sufficient to justify your inferring *Z* should you learn *Y*. In order to have probabilistic soundness you need *X & Y* to be highly probable. And although *X* is highly probable, it need not be the case that should you learn *Y*, *X & Y* would *then* still be highly probable. *Y* might, for instance, be just the sort of information that in the circumstances would undermine *X*. If we make the standard assumption often called 'Conditionalization' (see, e.g. R. C. Jeffrey, *Logic of Decision*, McGraw-Hill, 1965), *X & Y* would be highly probable should you learn *Y* if and only if $Pr(X \& Y/Y)$ is high. (What would happen should you learn *Y* and ... is a separate question because *Strengthening the Antecedent* fails for subjunctive conditionals.) The latter conditional probability simplifies to $Pr(X/Y)$.

The position then is that in a case where (i) '$X, Y, \therefore Z$' is valid and (ii) *X* is highly probable; in order for it to be warranted to infer *Z* should you learn *Y*, it must be the case that $Pr(X/Y)$ is high.

Let us now apply all this to whether the subjunctive phenomenon Anderson and Belnap point to counts against *EDS*. They present a case where someone, you, say, holds *A-or-B*, and is interested in inferring *B* on the addition of not-*A*. Thus we have in this case (i) that $A \vee B$ is highly probable, and if we assume *EDS* is valid, (ii) that '$A \vee B$, not-$A \therefore B$' is valid; and the question is whether you would be warranted in inferring *B* on learning not-*A*. The answer is that this will be warranted in the case described only if $Pr(A \vee B/\text{not-}A)$ is high. Now $Pr(A \vee B/\text{not-}A)$ simplifies by the calculus to $Pr(B/\text{not-}A)$, so the requirement is that the latter be high. By Conditionalization this obtains when should you learn not-*A*, *B* would then be highly probable, but in this case you will grant that should *A* turn out false, then *B* would most likely be true – *and this constitutes drawing a subjunctive conclusion of the kind in question.*

Therefore we have shown that if we make the assumption that *EDS* is valid together with assumptions like Conditionalization which are widely accepted and non-question-begging, the kind of subjunctive phenomenon Anderson and Belnap are referring to is exactly what we should expect, and so no evidence against *EDS* being valid.

Two objections might be raised. The first is that we have controversially identified holding a proposition with giving it high subjective probability. However, all our argument actually needs is that high subjective probability is a necessary condition for holding true, and that is non-controversial (by philosophical standards). The second is that Conditionalization is disputable. But it is plausible that any reasonable thesis about the impact of new evidence would enable a similar prediction of the subjunctive phenomenon by a believer in *EDS*.

Monash University
Clayton, Victoria, Australia 3168

©FRANK JACKSON *and*
LLOYD HUMBERSTONE 1982

NEGATIVE EXISTENTIAL PROPOSITIONS

By BARRY MILLER

A LARGE part of C. J. F. Williams' recent and impressive book *What is Existence?* is directed against what I shall call the two-sense thesis of '—— exists' – the view that '—— exists' does not always occur as a second-level predicate, but may sometimes be used as a first-level one. Early in the book he announces:

> If someone wishes to persuade me that '—— exists' can function as a first-level predicate, he will have to show me a sentence which I can recognize as usable outside philosophy and which is formed by wrapping '—— exists' round a genuine proper name. Until this is done, I protest that I simply do not understand what is meant by filling the gap in '—— exists' with a proper name. (p. 79)

He then goes on to consider various of the sentences that have been proposed as being not only obviously intelligible but as resulting 'from wrapping "—— exists" round a proper name'. (p. 79)

Among such sentences are 'Arkle no longer exists', 'Arkle came to exist', and 'Arkle ceased to exist', all of which he claims would not only generate a paradox if '—— exists' were a first-level predicate, but would do so for the same reason in each case. The reason is that an account of the truth-conditions of each sentence would have to include the sentence 'Arkle does not exist'. And the trouble with that, he argues, is that to treat '—— does not exist' as a first-level predicate in that sentence – which seemingly we should be bound to do if '—— exists' were a first-level predicable – would commit us to admitting that non-existence was a property that Arkle could acquire only when there was no longer an Arkle to acquire it. An attempt to avoid the paradox is examined and found wanting. He thus concludes that, their appearance notwithstanding, these sentences leave him unpersuaded as to the truth of the two-sense thesis.

I disagree with his conclusion because, as I hope to show, the attempt to avoid the paradox is lacking in the defects that Williams attributes to it. He rightly takes it to be based on the view that 'Arkle does not exist' is 'ambiguous in that negation can be regarded as either internal or external to the proposition as a whole'. However, as he understands it, the distinction between internal and external negation is to be equated with one 'between saying that non-existence is truly predicable of Arkle and saying that the proposition that Arkle exists is false. ... "Arkle no longer exists" is true, not because the predicate "—— does not exist" is now, though earlier it was not, predicable of Arkle, but because the proposition "Arkle exists" is now, though earlier it was not, false'. (p. 124)

Attributing this view to Barry Miller, he criticises it on two scores. Firstly, he argues that internal/external negation is 'a distinction without a difference'. Secondly, he seems to claim that, even were such a distinction allowed, supporters of the two-sense thesis would face the dilemma of having to deny either that 'properties are what predicables stand for, or that predicables are what are obtained by removing proper names from propositions'.

> Removing a proper name from a negative proposition automatically produces a predicable which stands for a negative property; and, equally, ascribing such a property to an object is done by attaching a negative predicable thus obtained to the name of that object and thus producing a

> negated proposition. Those who deny that properties are what predicables stand for, or that predicables are what are obtained by removing proper names from propositions, owe us alternative accounts of *property* and *predicable*. (p. 126)

The point of this paper is to take issue with both criticisms, and thereby to show that there are clear cases of 'obvious intelligible' existential propositions for which no paradox arises even if '—— exists' is taken to be a first-level predicable.

I

As for the first criticism, it rests on a misunderstanding of the distinction between internal and external negation. It is not, as Williams thinks, that 'It is not the case that (Arkle exists)' says something about a proposition whereas '(Arkle) (does not exist)' says something about an individual (p. 124). On the contrary, both say something about the individual Arkle; they differ only in what they say about him. The latter says that non-existence is had by Arkle; the former denies that existence is had by Arkle. That scarcely seems to be a distinction without a difference.

Confirming that view is the fact that '*X* is not moral' may mean either of two things. It may mean that *X* has the 'property' of being non-moral; alternatively, it may simply be denying that he has the 'property' of being moral. Internal negation [viz. '(*X*) (is not moral)'] is being used in the first case, but external negation [viz. 'It is not the case that (*X* is moral)'] in the second. If, therefore, the distinction between internal and external negation were one without a difference those two renderings should mean the same. Yet that is just what they do not mean, for the first is to be taken as '*X* is immoral' but the second as the quite different '*X* is immoral or amoral'. The distinction is therefore a substantive one. In denying this Williams attaches great weight to the absence of the distinction from classical logic. He does admit, however, that such a consideration need not be decisive; and indeed it clearly is not.

What then of his second point that, even if he were to allow the use of external negation, those supporting a first-level use of '—— exists' would still have to admit to a new property – a negative one at that – being ascribed to an Arkle no longer able to acquire new properties? Well, although one would not want to deny his claim that predicables are obtained by removing proper names from propositions, one might well want to question whether every predicable stands for a real property. Some, I suggest, might stand for what I shall call Cambridge properties.[1]

---

[1] Geach has coined the term 'Cambridge change' to distinguish changes like butter's rising in price from real changes like butter's melting. I am simply extending the application of the adjective 'Cambridge' from changes to properties. Being at two dollars a kilo would be a Cambridge property of butter, whereas being melted (i.e. liquid) would be a real property.

Despite the terminology, there is nothing strange about this notion. We have only to recognize that no real change need be produced in *Y* by his being famous, for he may not even be alive when it happens. No real change is produced in a *Y* who, with no change in his own height, becomes shorter than a brother whose height has increased. No real change is produced in *Y* by his becoming temporally prior to a newly born nephew when previously there had been no nephew to be prior to. No real change is produced in *Y* by his being held responsible (whether in his own lifetime or later) for an event that might be traced back to a much earlier action of his. In each of these cases a new predicate becomes true of *Y*, viz. '—— is famous', '—— is shorter than his brother', '—— is temporally prior to his nephew', '—— is held responsible for event *Z*'. In no case, however, does any new quality or characteristic come to be had by *Y*: that is why I say what he acquires are not real properties, but only Cambridge ones. Hence, although it may be true that to every predicate there is a corresponding property, it is not true that the corresponding property has always to be a real rather than a Cambridge one. A further example is obtained by removing the name 'Hitler' from the proposition 'If anyone loved Hitler, he was deluded' to obtain the predicable 'If anyone loved ——, he was deluded'. What kind of real property does that predicable stand for? Surely it is ontological licence to suggest that it stands for any.

Naturally, anyone denying that the predicable stands for a real property needs to give some sort of account of how to distinguish predicables which do stand for properties from those which do not. Let me say immediately that there are no purely logical criteria for determining which predicables stand for real properties. There are such criteria, however, for determining at least some of the predicables that stand for Cambridge properties. Such predicables are ones which, although they might be required for the drawing of inferences, have no part at all in what Dummett has called the constructional history of sentences.[2]

For example, 'If anyone loved ——, he was deluded' might have to be recognized as a predicable for the purpose of making certain inferences,[3] e.g.

> If anyone loved maniacs, he was deluded.
> Hitler was a maniac.
> Therefore, if anyone loved Hitler, he was deluded.

The same predicable, however, plays no part in the constructional history of 'If anyone loved Hitler, he was deluded', which goes as follows:

---

[2] The constructional history of a sentence is concerned not with the order in which it was formed by the speaker, but with the order in which its logical parts must be conceived as being put together if we want to establish the sentence's truth-conditions.

[3] This point is made by Michael Dummett in Ch. 2 of his *Frege, Philosophy of Language* and in Ch. 15 of his *The Interpretation of Frege's Philosophy*.

1. From the arbitrary propositions 'Gisela loved Hitler' and 'Tom was deluded' remove the proper names to form the predicables '—— loved Hitler' and '—— was deluded'.[4]
2. Insert the same name (e.g. 'Eva Braun') in each of the gaps to form two propositions 'Eva Braun loved Hitler' and 'Eva Braun was deluded'.
3. Operate upon those two propositions with the operator 'if —— then . . .' to form 'If Eva Braun loved Hitler, Eva Braun was deluded'.
4. Remove 'Eva Braun' from that proposition to form the predicable 'if —— loved Hitler, —— was deluded'.
5. Conjoin this predicable with the second-level predicable 'For anyone, he —— he ——' to form 'For anyone, if he loved Hitler, he was deluded'.

Since the first-level predicable 'If anyone loved ——, he was deluded' is absent from that constructional history it cannot stand for a real property. If it stands for any property at all, it can only be for a Cambridge one.

Not even all predicables that are part of the constructional history of a proposition stand for real properties. Those that would not are ones that include propositional operators, e.g. 'If —— loved Hitler, then —— was deluded'. Among others would be those of the form '—— is *F* or is . . . *G*' and '—— is *F* & . . . is *G*'. Even when the constituent predicables '—— loved Hitler', '—— was deluded', '—— is *F*' and '. . . is *G*' happen to stand for real properties, the composite predicables stand for no *real* property over and above those stood for by the constituent predicables.

That, I think, is as far as one can go in using purely logical criteria to distinguish Cambridge properties from real ones. So far as concerns the predicables which remain untouched by these logical criteria, we must have recourse to empirical considerations. What are they? One might say, correctly, that real properties are those that can be either acquired or lost in real changes; but that would only move the problem back a stage, since no hard and fast criteria have been produced for real changes. However, although empirical criteria for real properties are decidedly elusive, that is not to say that the distinction between real and Cambridge properties is illusory. Rather, it means only that for some properties it is impossible to say whether they are real or Cambridge ones. Yet, although borderline cases may be insoluble, others are quite clear-cut. For example, the properties corresponding to the predicates '—— is famous', '—— is shorter than his brother', and '—— is temporally prior to his nephew' are clearly not real but Cambridge, since even when these are true of *X*, his having the corresponding properties may make not the slightest difference to him. It is equally clear that it does make a difference to *X* that he have the properties corresponding to the predicables '—— is deluded', '—— loves Hitler', '—— is intelligent', or '—— is melted'. Such properties are real ones.

---

[4] A Tarskian account of the constructional history would of course begin not with atomic sentences, as a Fregean account does, but with atomic *open* sentences.

Having made these points it is now possible to comment on Williams' claim that 'removing a proper name from a negative proposition automatically produces a predicable which stands for a negative property'. Let us examine what this might mean for 'Arkle does not exist', considering it first as containing external negation and then as containing internal negation. In the external negation case removing 'Arkle' results in the predicable 'It is not the case that —— exists', which Williams would regard as standing for a negative property. It is clear from the foregoing, however, that if it really does stand for a negative property, the property can no more be a real one than is the property for which 'If anyone loves ——, he is deluded' stands. Why? Because the predicable has no part in the constructional history of 'It is not the case that (Arkle exists)'. Consequently, if 'It is not the case that —— exists' were to stand for any property of non-existence, that property could not be a real one. But, if it is not a real one there can be no paradox, since there is nothing whatever paradoxical about a deceased Arkle acquiring a *Cambridge* property. Paradox could arise not from non-existence being a property, but only from its being a real one – which it is not. So much for considering 'Arkle does not exist' as being negated externally.

As we shall now see, the paradox still would not arise even if, as Williams seems to think, the negation in 'Arkle does not exist' were to be internal rather than external. Of course, he is right in thinking that a predicable can be obtained by removing 'Arkle' from the proposition, though this time it will not be 'It is not the case that —— exists', but simply '—— does not exist'. The question is whether that predicable stands for a real rather than a Cambridge property, and the first thing is to decide whether it forms part of the sentence's constructional history, which is as follows:

1. Remove the proper name from an arbitrary sentence, 'Caesar exists' to obtain the predicable '—— exists'.
2. Negate this predicable to form '—— does not exist'.
3. Insert 'Arkle' in the gap in '—— does not exist' to form '(Arkle) (does not exist)'.

Since '—— does not exist' does form part of this constructional history, it is still an open question whether the property for which it stands is a real one. Indeed, some might argue that since supporters of the two-sense thesis regard existence as a real property they of all people would be bound to regard non-existence as real also. If existence were a real property, surely non-existence also would have to be real.

Underlying this suggestion is the view that an individual, *Y*, could lack some real property only by having another real property correlative to the one it lacks. Well, let us consider this suggestion, not in regard to existence but in regard to the property red. The question then is whether the absence of redness from something which could be red must bespeak the presence of a real property correlative to red. Certainly, if *Y* is a piece of wood then it can lack redness only if it has some colour or colours other than red, be it brown, cream, fawn or whatever – none of which could be dismissed as mere Cambridge properties. That does not settle the question, however, since the result is very different when *Y* is not a piece of wood but a piece of glass. Now, although glass is like wood in being something that *could* be red, it is also unlike wood in that its failure to be red does not mean that it is any colour at all: it may be quite colourless. To say that it is non-red, therefore, is not to say that it has any correlative property, or at least not any that is real. It might of course be said to have the property of being non-coloured; but then so too might a pain or a flash of insight, though their being non-coloured could hardly be claimed as a real, rather than a Cambridge property.

Reflecting on this example, it is not difficult to see that lack of a real property *F* bespeaks the presence of a correlative *real* property of non-*F* only if *F* and non-*F* are differentiae of the one generic property. Thus, *if* red and non-red were related as differentiae of the property being coloured, red could be lacking in a *Y* that was coloured only if *Y* had some *colour* other than red. The relevance to the discussion of non-existence is fairly clear. If lack of existence in *Y* (which had existed) were to bespeak the presence of non-existence as a real rather than as a Cambridge property, existence and non-existence should be related as differentiae of some generic property just as red and non-red would have had to be related to the real property of being coloured. For convenience let us call this generic property '*E*'. Then, just as red and non-red would have had to be understood as *coloured* red and *coloured* other than red, so existence and non-existence would have to be understood as being *E* in an existential way and being *E* in a non-existential way. Thus, whether *Y* existed or did not exist it would have some form of being: it would be *E*. But that is not only false, but is held to be so even by supporters of the two-sense thesis.

The property stood for by '—— does not exist' is therefore no more than a Cambridge one. Consequently, no paradox can arise from 'Arkle does not exist' – not even when negation is taken to be internal to it.

II

To my suggestion that no paradox would arise from 'Arkle does not exist' if it were treated as a case of propositional rather than of predicate negation, Williams' reply was firstly to deny that negation admitted of any such distinction and secondly to point out that, even if my analysis were correct, the proposition would still contain a first-level predicate (viz. 'it is not the case that —— exists') which presumably would stand for a negative existential property. The paradox would therefore still be with us unless we were to deny either

a. that predicables were formed by removing a proper name from a proposition
or
b. that predicables stood for properties.

I have argued that the distinction between internal and external negation is not a distinction without a difference, but is a substantive one. I have argued further that (a) and (b) do not exhaust the options for avoiding any paradox arising from 'Arkle does not exist'. I accept (a); and I accept (b) as well. What I do not accept is Williams' unspoken assumption, viz.

c. that any negative existential property must be a *real* property.

As I have shown, that is not true of the property stood for by the predicable 'it is not the case that —— exists'.

The same point holds even where the negation is internal and the predicable is '—— does not exist'. There too we have seen that the relevant property is a Cambridge one. Thus, no matter whether the distinction between internal and external negation is accepted or rejected, the result is the same. In neither case are we committed to Arkle's acquiring any property that is real. In neither case, therefore, does 'Arkle does not exist' generate the paradox which was thought to count against accepting '—— exists' as a first-level predicable.

Responding, therefore, to Williams' request to be shown a sentence which he 'can recognize as usable outside philosophy and which is formed by wrapping "—— exists" round genuine proper names', I once again propose 'Arkle no longer exists', 'Arkle came to exist' and 'Arkle ceased to exist'.

University of New England
Armidale, N.S.W. 2351, Australia

©Barry Miller 1982

REPLY TO MILLER

By C. J. F. WILLIAMS

BARRY MILLER rejects my arguments against the supposed distinction between internal and external negation in the case of singular propositions. He wishes, for instance, to distinguish '(*X*) (is not moral)' from 'It is not the case that (*X* is moral)', equating the former with '*X* is immoral' and the latter with '*X* is either immoral or amoral'. Miller presumably wishes to make the point that 'moral' and 'immoral' are not contradictories but contraries; but in that case they cannot provide examples of the supposed 'internal negation'. As I recognized in my book (see p. 45, n. 5), perfect contradictories are not easy to find. It is sometimes objected that, e.g. 'odd' and 'even' are not contradictories, because the Tower of London, say, is neither. In that case, '(The Tower of London) (is not odd)', as Miller would write it, is true. So, of course, is 'It is not the case that (the Tower of London is odd)'. Some, however, would baulk at asserting 'The Tower of London is not odd', and claim that it was not meaningful to predicate oddness or evenness or their negations of the Tower of London. If they were right, I would claim that 'It is not the case that (the Tower of London is odd)' is not meaningful either. A string of words does not become a significant proposition by being placed in argument position to a truth-function. This dilemma can be used to parry many attempts to force us to distinguish between internal and external negations of singular propositions.

Miller next introduces the contrast between 'real properties' and what Geach has called 'merely Cambridge properties'. Among his examples of the latter are those which the predicables '—— is famous' and '—— is held responsible for event *Z*' stand for. Such properties can, in a Pickwickian sense, be acquired by an object after it has ceased to exist. Sections 4 and 5 of Chapter V of my book are given over to discussion of just these sorts of case. But not all the examples Miller himself gives of 'merely Cambridge properties' are properties which an object could acquire after it has ceased to exist. We know from the *Theaetetus* that *being shorter than my brother* is a property I can acquire, not by shrinking, but by having a brother who grows while I stay still. This makes it a 'merely Cambridge property', but not one which I could acquire after I had ceased to exist. Miller is mistaken, therefore, in his claim (p. 186) that 'there is nothing whatever paradoxical about a deceased Arkle acquiring a *Cambridge* property'. So, for all Miller has said, paradox *could* arise from Arkle's acquiring the supposed property of nonexistence, even if his arguments for this being a merely Cambridge property were accepted. The major premiss for the rest of Miller's argument is that, for every ϕ, if ϕ is a merely Cambridge property, ϕ can be acquired by an object after the object has ceased to exist.

Since this is false, there is no need to examine the arguments in support of his minor premiss.

University of Bristol
Queens Road, Bristol BS8 1RJ

©C. J. F. Williams 1982

LOGICAL EQUIVALENTS AND LOGICAL FORM

By A. J. Dale

IN 'Action and Reaction', recently reprinted in his volume *Actions and Events* (1980, pp. 137-48), Davidson discusses and defends his notion of logical form. In particular he is concerned to differentiate between identity of logical form of two sentences on the one hand and merely logically equivalent sentences on the other. Not only does Davidson amplify his conception of logical form but he also gives examples to illustrate the difference between identity of logical form and logical equivalence. It is my purpose in this paper to show that the examples chosen by Davidson to illustrate this difference also show that his talk of 'logical form' is little more than a rhetorical device.

> To give the logical form of a sentence is to give its logical location in the totality of sentences, to describe it in a way that explicitly determines what sentences it entails and what sentences it is entailed by (op. cit. p. 140).

> We have given the logical form of a sentence when we have given the truth conditions of the sentence in the context of a theory of truth that applies to the language as a whole (op. cit. p. 143).

These two quotations crystallize the various formulations that occur throughout his writings. The connection between these two seemingly different explanations of what it is to give the logical form of a sentence is, of course, quantification theory. For quantification theory is amenable to a theory of truth which would satisfy Tarski's criteria, and translation of a sentence into quantificational form does explicitly determine its logical relations to the totality of sentences – or at least the totality of sentences in quantificational form.

As Davidson recognizes, this reference to quantification theory appears to introduce a certain relativity into the notion of logical form, for

> the location [of a sentence] must be given relative to a specific deductive theory; so logical form itself is relative to a theory. The relativity does not stop here either, since even given a theory of deduction there may be more

> than one total scheme for interpreting the sentences we are interested in and that preserves the pattern of entailments. The logical form of a particular sentence is, then, relative both to a theory of deduction and to some prior determinations as to how to render sentences in the language of the theory (op. cit. p. 140).

There is an ambiguity here, since there is the problem of identifying deductive theories. There are many different formulations of equivalent deductive theories and this fact has led to criticism of the attempt to reveal logical structure by translation into formal language equivalents of natural language sentences (e.g. Ziff *Understanding Understanding* 1972, p. 56ff.). Davidson's talk of the merits of *quantification theory* suggests, however, that it is not these different, equivalent formulations that he has in mind when he claims that logical form is relative to a background theory but rather the difference between any of these and some other non-equivalent deductive theory – an augmented or even non-standard theory, perhaps.

Be this as it may, I shall suppose that some particular formulation of standard quantification has been chosen. There now enters the second factor which contributes to the relativity of logical form. There are, as Davidson claims, 'more or less standard techniques for paraphrasing many sentences of natural language into quantificational languages' (Davidson, op. cit. p. 140): the fact that such standard techniques exist, he says, helps to conceal the relativity to a theory.

It is this second contributing factor which empties the notion of logical form of its content. Davidson considers the traditional example 'all whales are mammals': his claim is that anyone interested in its logical form would maintain that its form was that of a universally quantified conditional whose antecedent is an open sentence true of whales and whose consequent is an open sentence true of mammals. But this is not necessarily the case. It is possible to maintain that its form is of a negative existential conjunction one conjunct of which is an open sentence true of whales and the other conjunct an open sentence not true of mammals. Choosing the latter form may even be preferable for certain purposes – weaning students from Venn diagrams to quantification theory, for example.

A less worn example would be of a sentence containing 'any'. What is the logical form of 'if anyone leaves the meeting will be inquorate'? Is it a conditional, whose antecedent is an existentially quantified open sentence or, on the other hand, a universally quantified conditional? The two possibilities are, of course, logically equivalent within the chosen deductive theory. But how to distinguish which is to represent the logical form of the sentence in question? There is extra bite to this example, since it is possible to use this sentence to illustrate that the two possibilities are indeed logically equivalent and thus overcome the reluctance of beginners to accept the more puzzling of the rules of passage for quantifiers.

The same problem arose for Moore when he discussed Russell's definition(s) of definite descriptions, for, as Moore saw, Russell gives two seemingly incompatible accounts of what a sentence containing a definite description means (G. E. Moore, 'Russell's Theory of Descriptions', *The Philosophy of Bertrand Russell*, ed. P. Schilpp 1944, p. 221). On the one hand such a sentence seems to express a proposition which is a conjunction of three independent propositions and, on the other hand, it seems to express an existential proposition.

The problem, in one guise or another, is not new. What is curious is Davidson's seeming disregard of it. It may be noticed that in the three examples so far given in this paper, the two possible interpretations for each case would certainly satisfy his conditions for giving the logical form of the natural language sentences. In each case the possible interpretations explicitly determine which sentences they entail and are entailed by. In each case a theory of truth for quantification theory would provide their truth conditions. It would seem that in each case either interpretation could be regarded as giving the logical form of the original natural language sentence, even after limiting the background theory not just to standard quantification theory but to some specific formulation of it.

Since Davidson explicitly mentions standard techniques for paraphrasing into quantification theory and that these techniques help to conceal the relativity of logical form, the conclusion seems inescapable that for Davidson *any* of the possible interpretations would do. But then the claim that such an interpretation gives the logical form of the original sentence looks empty indeed. For, suppose we agree to paraphrase the sentence 'all whales are mammals' in the manner Davidson suggests that those with an interest in logical form would so paraphrase it, as a universally quantified conditional – '(x) (x is a whale → x is a mammal)', say. Then let us mention the relativity explicitly. Davidson's claim that the latter gives the logical form of the former amounts to this: relative to the paraphrase of 'all Fs are Gs' into the chosen formulation of quantification theory as '(x) (Fx → Gx)' the logical form of 'all whales are mammals' is '(x) (x is a whale → x is a mammal)'. Put like this, claims to give the logical form can be seen to be exceedingly uninteresting. It would seem that Davidson's use of the term 'logical form' was intended to add some weight to his discussions of action sentences and causal sentences since the term itself has a philosophic, if controversial, history. The attempt to relativize it removes its interest.

It remains an option to allow any and all of the possible logically equivalent sentences to represent the logical form of the natural language sentence – indeed, by identifying logically equivalent propositions Wittgenstein maintained a position tantamount to this. But not only does this lead to the perhaps unacceptable conclusion that the logical form of 'all whales are mammals' could be '$(x)\,(x \text{ is a whale} \rightarrow x \text{ is a mammal}) \cdot (2+2=4 \vee {\sim}\, 2+2=4)$' but it would also contradict Davidson's explicit example of difference of logical form.

The example of a pair of logically equivalent sentences that are not of the same logical form is a pair consisting of a conjunction and the conjunction formed from it by reversing the order of its conjuncts. In this case, Davidson claims, a theory of truth would provide the truth conditions of each conjunction and we could prove that one is true if and only if the other is. This is certainly the case and shows that the two sentences are logically equivalent but it does not show that they have different logical forms. Indeed both sentences satisfy the explicit criteria for identity of logical form quoted at the beginning of this paper – explicit determination of logical relations and identity of truth conditions. This will be the case with any pair of logically equivalent sentences formulated within quantification theory. In order to separate such a pair some other condition must be provided in addition to those given by Davidson.

The particular example of the two conjunctive sentences seems a particularly inappropriate illustration of difference of logical form. It is true that the two sentences are different sentences since their conjuncts are in reverse order but this fact is insufficient to guarantee difference of logical form. The major interest in logical form lies in the fact that different sentences may have the same logical form. It is Davidson's claim that one sentence can give the logical form of a different sentence. In the case of the two conjunctions, however, there seems no ground other than that they are different sentences for claiming that they are of different logical form. On any reasonable theory of logical form it should be the case that the two different conditional sentences 'if it rains the match will be cancelled' and 'the match will be cancelled if it rains' have the same logical form. It is of no matter whether the antecedent of the conditional precedes or succeeds the consequent in a conditional sentence: reasons of style or emphasis will govern the choice. It remains true nevertheless that either the antecedent or the consequent must be placed first in the sentence if we are to write in a linear script or talk in temporal order. Precisely the same factors should be taken into consideration in the case of conjunctions. The exigency of communicating by means of some linear order will force us to place one conjunct before the other just as, in the case of the conditional, it forces us to place either the antecedent or the consequent first. The only difference between the two cases is that whereas the clause beginning with 'if' can be moved to the front of the sentence to form a new syntactically correct sentence the clause beginning with 'and' cannot be so moved. But this would seem to be no more than a parochial trait of the English language. To dispel remaining doubts it would be better perhaps to consider the case of disjunction which is entirely parallel to the case of conjunction and the same considerations apply. The Latin constructions 'vel...vel...' and 'et...et...' show that it is possible to have constructions which do not retain this feature of English.

The preceding paragraph presupposes a 'reasonable' theory of logical form, but it may be the case that Davidson believes that different sentences in quantificational form should have different logical forms. This would create tension between the tolerance of paraphrase into quantificational form – the relativity of logical form to standard techniques for such paraphrasing – and the demand that different sentences within quantification theory have different logical form. But such a thesis is by no means in keeping with any acceptable logical form, for then we would have to maintain that '(x) (x is a whale → x is a mammal)' has a different logical form from '(x') $(x'$ is a whale → x' is a mammal)' since there is a different variable in each case. Yet, for all that Davidson has explicitly said on difference of logical form, there is nothing to rule out this case. For, under a suitable theory of truth we would be provided with the truth conditions of each and we could show that one was true if and only if the other is. But such a theory, which separated two sentences into different logical forms by the mere occurrence of a different variable, would not be worthy of the title 'theory of logical form'.

There are, of course, different ways to augment Davidson's notion of logical form with additional constraints but it is not apparent that they would prove acceptable to Davidson (e.g. G. Harman, 'Logical Form', *Foundations of Language* 1972, pp. 289–307). But, more to the point, it would seem unnecessary to his major theses to bring in the notion of logical form at all. It would seem that its presence in his seminal papers on action sentences and causal sentences was to place his work in a philosophical tradition, but it plays no working role in the papers themselves. If this claim seems extravagant the reader should compare 'On Saying That' with his recent paper 'Quotation' (*Theory and Decision* 1979, pp. 27–40): in the latter paper a theory of quotation is presented along the lines of the theory of indirect discourse presented in the former. The difference is that in his theory of quotation there is no mention of logical form whatsoever yet his argument loses none of its strength by its absence.

The University
Hull HU6 7RX

© A. J. DALE 1982

TOKEN RESISTANCE

By PETER M. SIMONS

THE arguments of Hugly and Sayward [1] against muddled thinking about the token/type distinction are welcome, but their own positive suggestions contain subtler errors. I am not convinced that the distinction is not an irretrievable muddle, but whether it is or not, any account of types will have to be more careful than those offered hitherto, and it is worth considering what obstacles such a theory has to surmount.

Hugly and Sayward assume, as do many, that the existence of types presents no problem. To the nominalistically inclined, this is questionable. Certainly, proofs that a certain language contains infinitely many expressions cannot relate to expression-tokens, and formal languages are usually taken to be at least denumerably infinite. Nevertheless, such proofs do not *exhibit* tokens of infinitely many types, any more than Euclid's proof that there are infinitely many primes exhibits expressions for infinitely many. What is given is a rule or recipe for constructing new expressions (or designating a new prime number) provably different from all those constructed up to a given point, thus reducing the assumption that there is a finite bound to absurdity. All this spills only finitely much ink. Although hardly popular, a strictly nominalistic approach to logic and language may not be unworkable. Such an approach was attempted by Leśniewski, and accounts for some of the difficulty experienced by logicians trained in the orthodox – and platonistic – tradition in understanding Leśniewski, whose systems are envisaged as growing in time by the addition, according to the directives, of new theses. (Cf. [2].) Such talk often meets with incredulity or downright ridicule, but is simply the consequence of a consistently nominalistic attitude to language. Moreover while the metalogical concepts of such an approach are somewhat different from usual, its actual proofs, both logical and metalogical, need differ little in structure or power from those of a platonistically-conceived system.

It is indeed surprising, if Hugly and Sayward are right that even noted authorities are confused about types and tokens, that generations of logicians *have* muddled through without serious difficulty. This suggests that the distinction is not crucial to the practice of logic and linguistics: we manage expressions without it. I think this clarity comes almost exclusively from our ability to manage and think about tokens, and that talk about types is largely carried over analogously from talk about tokens. Nor, as we shall see, is the analogy very exact.

The distinction exists chiefly to rationalize talk about different instances or occurrences of the same expression, and so, from a nominalistic point of view, was conceived in sin. However, even for this limited purpose, the distinction is not as clear and helpful as it may seem. A simple example will make this clear. Consider a simple language, say first-order, in which are included the names '*a*', '*b*', the predicates '*F*', '*G*', and the formulae (sentences) '*Faa*', '(*Faa* & *Gab*) ⊃ *Faa*'. We use here the convention that quotation-mark expressions name types, not tokens. Now one difference between '*Faa*' and '*Gab*' seems to be that the former contains two occurrences of the name '*a*', while the latter contains single occurrences of the two different names '*a*' and '*b*'. However, on closer reflection this cannot be right. '*Faa*' and '*a*' are types, and as such unique abstract individuals, standing over against the multiplicity of their respective tokens. If '*a*' occurs twice in '*Faa*', then either '*Faa*' contains two tokens, in which case it is an impossible hybrid, or there are several abstract '*a*'s, which are then presumably tokens of a higher, second-order type. But then we would not know which of these equiform types a token was a token of. And the story would not end there, because in '(*Faa* & *Gab*) ⊃ *Faa*' the type '*Faa*' itself would have to occur twice, so we need third-order types, and so on. The regress thus started is both uneconomical and vicious, because there is no point at which we reach unique types which account for the multiplicity of like tokens.

An account of the occurrence of type expressions within type expressions which avoids this regress can be found in Quine's *Mathematical Logic*. Quine writes

> what is an occurrence? what kind of thing, e.g. is the third occurrence of an expression x in an expression y, as distinct from the mere expression or typographic shape x itself? An answer, artificial but convenient and adequate, is this: an occurrence of x in y is an initial segment of y ending in x. ([3], p. 297.)

So, for instance, in '*Faa*' the first occurrence of '*a*' is '*Fa*', the second occurrence of '*a*' is '*Faa*'. The artifice of this device consists partly in its conventionality: a final segment of y beginning with x would have done just as well. The suggestion nevertheless has the merit, measured against intuitive expectations, that it makes the occurrence of a type expression itself a type expression, and yet manages to differentiate the different occurrences of a type expression within another. On the other hand, measured by these same expectations, it fails to capture the qualitative equality of the different occurrences of one and the same expression. Intuitively, Quine seems to be identifying occurrences with something else, namely initial segments. This would be more excusable if there were a one-to-one correlation between occurrences and initial segments, but there is not. Consider the situation among tokens. Here we can discriminate not only (tokens which are) occurrences of simple expressions, but also those which are occurrences of complex expressions. For example in the token

Fab ⊃ *Fab*

we can discriminate two occurrences of '*Fab*' as well as two occurrences of '*b*', and no occurrence of '*Fab*' is (obviously) an occurrence of '*b*'. We should expect a theory of the occurrence of types within types to do as much, but Quine's account cannot distinguish between occurrences of '*Fab*' and of '*b*', since they are conventionally identified with the same initial segments. Thus Quine's claim to adequacy is forfeit. A similar problem besets the symmetrical end-segment convention.

Could we somehow combine both conventions and home in on the wanted occurrence? Suppose we take an occurrence to be an ordered pair $\langle i, f \rangle$ consisting of an initial segment i and a final segment f? In fact this only suffices when we stipulate which expression they are segments of: the pair $\langle$'*Faa*', '*aab*'$\rangle$ could represent the occurrence of '*aa*' in '*Faab*' or the second occurrence of '*a*' in '*Faaab*', so we need an ordered triple $\langle e, i, f \rangle$, where e is the expression of which i is an initial, f a final segment.

While such a convention provides us, under suitable assumptions about the existence of ordered triples, with abstract objects in one-to-one correlation with occurrences of expressions, these objects are themselves no longer expressions, as we should expect, and as Quine's convention at least allowed. Further, the qualitative equality of different occurrences of the same expression remains unrespected, as it does by Quine. Simple token expressions are the literal *parts* of the complex token expressions they build up, and complex tokens may be part of larger complex tokens. Tokens – whether simple or complex – of one type are, within the bounds laid down by the conventions of writing, alike. In the simplest imaginable case this likeness consts solely in physical similarity. The difficulties of both the naive theory, which ends in regress, and Quine's theory, or the modified version just mooted, strongly suggest that not all of the essential properties of and relations among tokens can be carried over analogously to types, and therefore that a theory of types has little explanatory power. This is not to deny that we can and do use singular terms like 'the expression "*Faa*"' perfectly happily, nor indeed that such terms denote abstract objects. Rather it suggests that the way to understand clearly what a predication about types is saying is to trace it back to a sister-predication about tokens. For instance, the problematical sentence 'The formula "*Faa*" contains two occurrences of the name "*a*"' may be traded in for the sentence 'Any formula equiform with "*Faa*" contains two names equiform with "*a*"', where now the inner quotes form a name of the token contained within them (cf. [4], p. 284f.), and the terms 'formula' and 'name' relate to tokens. It would be rash to declare the two sentences synonymous, but the second has the merit of using the term 'contain' in its usual mereological sense. Further, the second sentence opens the way to recognizing something which the type analysis tends to cover up, namely the variety and partly conventional nature of equiformity. For example, for spelling purposes the tokens 'a' and '*a*' count as uniform, but for typesetting purposes they do not.

Hugly and Sayward consider two alternative ways to view types: as patterns or as classes. They prefer the class view. I prefer the pattern view. They give at least two reasons for their preference. If patterns are always exemplified patterns, then the pattern view is inadequate, since only finitely many patterns of an infinite language are exemplified. ([1], p. 182). A platonist need not baulk at unexemplified patterns, however, nor should this disturb the logician or linguist, provided the *basic* expressions are exemplified (so we know which they are) and we have effective rules for constructing a token of any hitherto unexemplified type. A weaker view would perhaps regard all patterns as exemplified, but talk about unexemplified patterns as disguised modal talk about what could be the case. However, Hugly and Sayward require that the metalanguage of a first-order theory be extensional ([1], p. 184) – though they do not give reasons why – so they would reject any modal analysis. They also object that the notion of a pattern is unclear. (*Ibid.*) That may be so, but if we are trading uncertainties, I think the notion of an abstract class is unclear (cf. [5]), so the objection cuts both ways.

Their simple first-order language consists of basic expressions, which are classes of tokens similar to those printed, and complex expressions are construed as various kinds of ordered *n*-tuples which render in set-theoretic terms both the ordering and hierarchical aspects of sentence construction. Such *n*-tuples may be arbitrarily complex, but the recursive rules allow an effective decision procedure as to whether a given set is an expression or not. The basic expressions are effectively equivalence classes of tokens under the relation of equiformity or physical similarity. This is intended to secure simultaneously the initial connection between tokens and types, and the abstract status of types. But types, as abstract objects, are supposedly immune from the contingencies affecting their tokens. Yet the upshot of the Hugly-Sayward definition is that an expression such as 'x', being a class, is determined by which tokens actually exist. Had some of these not existed, say if the printer had run off one less copy, or had others existed, then the *type* 'x' would have been different from what it actually is. This argument is a general one against construing abstract objects 'understood through' equivalence relations as equivalence classes, where classes are taken as determined by their members (cf. [6]). Hugly and Sayward might counter that they are only interested in an extensional account of types, but since the intuitive concept of a type is not extensional, they are then offering an alternative rather than an analysis.

For the purposes of explaining the structure of complex expressions, it does not much matter on their account what the basic expressions are. They could just as well have been the first five numbers, or any other five abstract objects. Since their account of the relation between tokens and types is very liberal as to what can count as tokens ([1], pp. 186–7) the initial link is arbitrary. Given any system of abstract expressions, there are many ways in which this can be realized or tokened. Not only can we take anything we like as tokens of the primitive signs (provided only that they are different from one another and from tokened complex expressions), but also the forms of combination corresponding to the set-theoretic structure of complex expressions may vary widely. For instance, a first-order language could use Peano–Russell, or Lukasiewicz notation, or Frege's two-dimensional *Begriffsschrift*. These are only, for Hugly and Sayward, different possible notation or tokening systems for realizing one and the same abstract expressions. They require only that each token represent a unique expression ([1], p. 186). In fact the natural tokening system they offer falls short of this requirement, since the token

N P′x′′P′x′

is ambiguous as to whether it tokens ⟨D, ⟨⟨B, C⟩, ⟨A, C⟩⟩, ⟨⟨C, ⟨B, C⟩⟩, ⟨A, C⟩⟩⟩, i.e. is to be read as N(P′x′)(′P′x′), or tokens ⟨D, ⟨⟨B, C⟩, ⟨⟨A, C⟩, C⟩⟩, ⟨⟨B, C⟩, ⟨A, C⟩⟩⟩, i.e. is to be read as N(P′x″)(P′x′), A minor adjustment will right this, but it does indicate the possibility of ambiguity, which is realized frequently in natural languages.

It is usual to regard first-order languages formulated with different primitive signs (indicated by the written tokens) and different notational conventions as different languages. These languages may indeed be isomorphic, and the more abstract structure revealed thereby is nearer to the kind of structure Hugly and Sayward have in mind. Their structures of types are too abstract to be expressions of a language. They may object that the choice of certain classes as primitive expressions suffices to individuate the language. But it does not. Consider an imaginary language, Reverse-English, in which the basic expressions (be they words or morphemes) are the same as in English, but the word-order is the exact reverse of English. (Reverse-English of sentence a be would this then.) Reverse-English is isomorphic to English, and has the same primitive expressions, so by the Hugly–Sayward criterion would be the same language as English, which it is not. Consistently with their view, they take expressions of Morse Code (as used in sending English messages) as expressions of English on a par with any other: they simply belong to a different tokening system. By this criterion, written English, shorthand, impulses on magnetic tape, Gödel numbers and much else count as expressions of English. This ignores the priority of the spoken language. Dots and dashes of Morse Code are not expressions of English: they are *encodings* of the letters and punctuation marks used in writing English. Even the words of written English are expressions of English at best in a secondary sense. They too are encodings (albeit so familiar we forget this) of the spoken language. The various encodings of the spoken language stand in varied and highly complex relations to it and to one another. They have their own rules and peculiarities, and to a greater or lesser extent a life of their own. This makes it doubtful that the ascetic set-theoretic means at the disposal of Hugly and Sayward could accommodate the variety of combinatory and realizatory relationships present in natural languages and their encodings. Hence they consign such relationships to systems of tokening. Their relative success with a simple first-order language is deceptive, since, firstly, such languages are, unlike natural languages, primarily written, and secondly, lack the multiple articulation of natural languages. The problem with applying their analysis to natural languages lies less with the recursive specification of the expressions of the language ([1], p. 187), than with the fact that the type/token distinction applies to natural languages at *each level* of articulation anew: be it phonological, graphological, morphological or semantic. The former difficulty is a technical one, but the latter is a defect in principle. Similarly, Morse Code or any other encoding system is subject to the distinction independently of its use for encoding this or that natural language.

One reason why I prefer the pattern approach to types is that we can be sure that this applies at each level of articulation of a language. The patterns are more closely determined by the nature of their respective tokens. For written languages the graphological patterns are patterns of marks, for spoken they are patterns of certain kinds of sound, for Braille patterns of dots on paper, and so on. Another reason for preferring patterns is that there is a use of patterns in logic which is crucial to its employment, yet which is not covered by the Hugly–Sayward approach. A single expression may exemplify several different patterns, and different expressions may have some, but not all patterns in common. For instance, both 'Fxy' and '$(z)(Fxz \supset Gyz)$' exemplify the pattern of being a formula with only the two free variables 'x' and 'y', but one is atomic while the other is a formal implication. Again, 'Faa' exemplifies numerous patterns, some of which we shall presently notate. While all the information about the patterns exemplified by a given formula or other expression can be gathered from a perusal of the set-theoretical structures Hugly and Sayward take as the types of such expressions, this amounts only to discerning patterns among these structures – or perhaps more probably, patterns among the expressions designating them. These other patterns are not themselves represented, and the rules for defining the structures do not help us to understand how we find such patterns, which they ought to if they are to offer an explanation of types and not merely a model. Set-theoretical models are easy enough to construct, but display certain aspects of arbitrariness, for instance in the definition of ordered pair (cf. the timely criticism of the set-theoretical model of ordered pairs in [7]). Since any model is as good as any other, we could have chosen to model a language in the natural numbers, using Gödelization.

One hurdle to be overcome in developing an analysis of types along these lines is the objection that patterns, unlike the expressions which exemplify them, cannot be given names, or more generally, designated by singular terms, since patterns are, or are like, Fregean functions: they are *ungesättigt*, and cannot, like Fregean objects, be named. (Cf. [8], pp. 146–8; [9], p. 142, pp. 148–9.)

Whatever may be the case for functions in general, the patterns that interest us here can be named. We need only develop a suitable notation, and notation for patterns exists already. For example, the rhythm of lines of poetry is a pattern in the sound material, and we have the notation '˘', '¯' for distinguishing these rhythms. A suitable beginning for a notation for patterns in a first-order language can be made using Frege's device of Greek place-holder letters (cf. this use in [9], p. 146.) We sketch how this goes: adopt Greek letters *ξ*, *η* as place-holders for names; *φ*, *ψ* as place-holders for predicate letters. Single quotes are used to designate types, double quotes to designate patterns. For example, "*Fξη*" names the pattern, common to '*Fab*', '*Fba*', '*Faa*' etc., of being an atomic formula beginning with '*F*' and followed by two tokens of (not necessarily distinct) names. "*Fξξ*" names the pattern, common to '*Faa*', '*Fbb*', etc., of being an atomic formula like that above with the extra stipulation that a single name occur twice. Then "*Fξη*" = "*Fηξ*" ≠ "*Fξξ*" = "*Fηη*". Though "*Fξξ*" and "*Fξη*" are different patterns, whatever exemplifies the first exemplifies the second. We can now say that any token of '*Faa*' exemplifies all of "*Faa*", "*Faξ*", "*Fξa*", "*Fξξ*", "*Fξη*", "*φaa*", "*φaξ*", "*φξa*", "*φξξ*", "*φξη*". With a richer notation we could name further patterns. Notice that as a limiting case, a fully materially determined pattern is just a type: "*Faa*" = '*Faa*'. Now it is clear that no taoken of '*Faa*' contains any token of '*Fξη*' as part. But is it so obvious that it does not contain a token of "*Fξη*"? Obviously we cannot *isolate* a pattern, whether between quotes or any other way, from all material whatsoever. The best we can do is exhibit (here between quotes) something which exemplifies the pattern. We could choose a token of '*Faa*' itself, but since this exemplifies many patterns, it is not then clear which one is being demonstratively exhibited. Hence our convention. Here no pattern is exemplified in isolation: the Greek letters, or something like them, are indispensable. Tokens of patterns are, unlike the expressions they are found in, *ungesättigt*, or as Frege also less metaphorically put it, *ergänzungsbedürftig* – in need of supplementation – or *unselbständig* – incapable of independent existence. They may no more be isolated between quotes than a Cheshire cat's smile could really remain in the absence of its bearer.

Now, however, the objection may be raised that one cannot talk about *tokens* or *instances* of the sorts of pattern mentioned. The only tokens, it may be said, are the complete expressions themselves, and the patterns these exemplify are nothing other than various common properties of the expressions. (Cf. [10], pp. 224–5.)

Even if we can present in context what cannot be presented in isolation, this must be a universal property. If that is right, then the view that, as patterns, types are abstractible from a nominalistically acceptable base of tokens, cannot be upheld. (On the notion of 'base' and the kind of abstraction here envisaged, cf. [6].) Either the base is nominalistically acceptable, and fails to provide material for the abstraction of patterns, or it already embraces patterns, in which case it is platonistic from the start.

There is a way between the horns of this dilemma if one recognizes the existence of *particulars* which are *unselbständig,* such as smiles, cricket strokes ('a fine example of the cover drive') and – tokens of patterns. On this view, every token of '*Faa*' includes a token of "*F*ξξ", though not separably. Such dependent particulars, which, following Husserl ([11]), I call 'moments', are familiar to historians of philosophy as accidents or modes, and can be put to a variety of uses, not least the elucidation of Frege's obscure metaphor of *Ungesättigtheit* (cf. [12].) Since many moments are perceivable (cf. [13], §4), they are epistemologically relatively unproblematic and nominalistically acceptable. So a nominalistic approach to language and logic need not forego patterns: it can treat them, as it treats expressions, as tokens. A theory of types seeing them as abstracted from an ontological base of dependent and independent particulars would then have a richer base from which to start, but the structure of types would also be to a greater extent a mere reflection of the structure of tokens, reducing considerably the explanatory role of types.[1]

Universität Salzburg
Institut für Philosophie
Franziskanergasse 1
A-5020 Salzburg, Austria

© PETER M. SIMONS 1982

REFERENCES

[1] P. Hugly and C. Sayward, 'Expressions and Tokens', *Analysis*, 41.4 (1981), 181-7.
[2] P. M. Simons, 'On Understanding Leśniewski', *History and Philosophy of Logic*, 4 (1982), 1-27.
[3] W. V. O. Quine, *Mathematical Logic*, rev. ed., New York: Harper & Row, 1951.
[4] H. Reichenbach, *Elements of Symbolic Logic*, London: Macmillan, 1947.
[5] P. M. Simons, 'Plural Reference and Set Theory', in B. Smith, ed., *Parts and Moments*, Munich: Philosophia, 1982, pp. 199-256.
[6] ——, 'Abstraction and Abstract Objects', in E. Morscher, O. Neumaier and G. Zecha, eds., *Essays in Scientific Philosophy*, Bad Reichenhall: Comes, 1981, pp. 355-70.
[7] H. Hochberg, 'The Winer-Kuratowski Procedure and the Analysis of Order', *Analysis*, 41.4 (1981), 161-3.
[8] P. T. Geach, 'Frege', in G. E. M. Anscombe and P. T. Geach, *Three Philosophers*, Oxford: Blackwell, 1961.

---

[1] I wish to thank the referee and Wolfgang Künne for their help. An earlier version was given at the University of Manchester in February 1982.

[9] ——, 'Names and Identity', in S. Guttenplan, ed., *Mind and Language*, Oxford: Clarendon, 1975.

[10] ——, 'Quine on Classes and Properties', in *Logic Matters*, Oxford: Blackwell, 1972.

[11] E. Husserl, 'Zur Theorie von den Ganzen und Teilen' ('On the Theory of Wholes and Parts'), in *Logische Untersuchungen*, 5th ed., Tübingen: Niemeyer, 1968, Vol. II/1 (*Logical Investigations*, tr. J. N. Findlay, London: Routledge & Kegan Paul, 1970, Vol. II.)

[12] P. M. Simons, 'Unsaturatedness', Grazer Philosophische Studien 14 (1981), pp. 73-96.

[13] K. Mulligan, P. M. Simons and B. Smith, 'Truth-Makers', Philosophy and phenomenological Research forthcoming.

INTUITIONISM DISPROVED?

By T. WILLIAMSON

PERENNIAL philosophers' hopes are unlikely victims of swift, natural deduction. Yet anti-realism has been thought one. Not hoping for anti-realism myself I here show it, lest it be underestimated, to survive the following argument, adapted from W. D. Hart ('Access and Inference', *Proceedings of the Aristotelian Society*, sup. vol. LIII (1979) pp. 156, 164-5; he credits first publication to Fitch).

Call (1) 'strong anti-realism' the view that all truths are known and (2) 'weak anti-realism' the view that all truths are knowable. In formal schemata with obvious notation:

(1) $P \rightarrow KP$

(2) $P \rightarrow \Diamond KP$

Strong anti-realism is *obviously* false: the number of books in my room may be odd even if I never bother to count them. Only weak anti-realism can be taken seriously. Unfortunately, (2) entails (1): since strong anti-realism has been rejected, so should weak anti-realism be, and any form of verificationism or idealism that entails it. The proof requires a minimum of modal and epistemic logic: the necessitation rule – if $\vdash P$ then $\vdash \Box P$ – the definition of possibility in terms of necessity – $\Diamond P$ as $\sim\Box\sim P$ – (though not, as Hart seems to think (p. 156), the rules that necessity entails truth and that the consequences of necessities are necessary) and two trite principles about knowledge – knowledge entails truth, and knowledge of a conjunction entails knowledge of its conjuncts:

(3) $KP \rightarrow P$

(4) $K(P \,\&\, Q) \rightarrow (KP \,\&\, KQ)$

Now assume:

(5) $P \,\&\, \sim KP$

By an instance of (2):

(6) $\Diamond K(P \mathbin{\&} {\sim}KP)$

By definition of $\Diamond$:

(7) ${\sim}\Box{\sim}K(P \mathbin{\&} {\sim}KP)$

Now assume:

(8) $K(P \mathbin{\&} {\sim}KP)$

By an instance of (3):

(9) $P \mathbin{\&} {\sim}KP$

From (8) by an instance of (4):

(10) KP

The second conjunct of (9) contradicts (10), so their assumption (8) may be negated:

(11) ${\sim}K(P \mathbin{\&} {\sim}KP)$

By necessitation (since (11) has no assumptions):

(12) $\Box{\sim}K(P \mathbin{\&} {\sim}KP)$

(7) contradicts (12), so (7)'s assumption (5) may be negated:

(13) ${\sim}(P \mathbin{\&} {\sim}KP)$

But (13) is equivalent to (1). Hence (2) entails (1) and anti-realism is untenable.

I shall not challenge the principles of modal and epistemic logic used (quibbles aside): they seem reasonable. However, the argument fails against the most significant current form of anti-realism: intuitionism.

How should 'KP' be interpreted? That P is known now? Or that it is, has been or will be known at some time or other? Consider the first alternative. If 'K' means present knowledge throughout, the argument remains valid. But then (2) says that whatever is true is knowable *now*, which is no longer integral to anti-realism. Why ever should anti-realists not allow some truths knowable only after the event? They could even give $P \mathbin{\&} {\sim}KP$ as an example. They need insist only that all truths be knowable at *some* time to remain anti-realists (perhaps that time must be future, not past: I ignore this inessential complication). So they are committed only to:

(2′) $P \rightarrow \Diamond\exists t\, K_tP$

where 'K_tP' means that P is known at time t. All knowledge statements must now be tagged for time (I use quantifiers and variables for clarity: operators would do). Then, where n is the present, (5) becomes:

(5′) $P \mathbin{\&} {\sim}K_nP$

An instance of (2′) now gives:

(6′) $\Diamond\exists t\, K_t(P \,\&\, {\sim}K_n P)$

which converts to:

(7′) ${\sim}\Box{\sim}\exists t\, K_t(P \,\&\, {\sim}K_n P)$

To contradict (7′), one must refute:

(8′) $\exists t\, K_t(P \,\&\, {\sim}K_n P)$

Subscripts turn (3) and (4) into (3′) and (4′):

(3′) $K_t P \rightarrow P$

(4′) $K_t(P \,\&\, Q) \rightarrow (K_t P \,\&\, K_t Q)$

(4′) needs the same subscript throughout if knowledge of a conjunction is *ipso facto* knowledge of its conjuncts; I shall not use the time needed to infer between conjunctions and conjuncts to undermine the argument. An instance of (3′) and (8′) give:

(9′) $P \,\&\, {\sim}K_n P$

But an instance of (4′) and (8′) give:

(10′) $\exists t\, K_t P$

and the second conjunct of (9′) does not contradict (10′). In fact, (8′) cannot reduce to absurdity: I verified it by noticing, five minutes later, that at the time to which the 'n' I wrote referred my book was open at page 156, but that at the time I did not know this. The argument cannot run on these lines.

'~KP' in (5) must be understood as incompatible with (10′), whose interpretation the requirements of anti-realism fixed: for the '~KP' in (9) 'comes from' (5). Hence '~KP' must be at least as strong as ${\sim}\exists t\, K_t P$, hence 'KP' at least as weak as $\exists t\, K_t P$: but it can hardly be weaker. So 'KP' must mean $\exists t\, K_t P$ in (5): no compromise between the present and eternity will do. 'K' is uniformly interpretable throughout the proof, which should succeed as at first.

To challenge its success intuitionist logic must now be involved. The argument as far as (13) may be intuitionistically valid: intuitionists might reject the definition of '$\Diamond P$' as ${\sim}\Box{\sim}P$, as they reject that of '$\exists x\, Fx$' as ${\sim}\forall x{\sim}Fx$, but could still accept the inference from $\Diamond P$ to ${\sim}\Box{\sim}P$, as they accept that from $\exists x\, Fx$ to ${\sim}\forall x{\sim}Fx$. However, intuitionistically (13) does not entail (1). ${\sim}(P \,\&\, {\sim}KP)$ entails only $P \rightarrow {\sim}{\sim}KP$; as 'K' is now understood:

(14) $P \rightarrow {\sim}{\sim}\exists t\, K_t P$

and the elimination of double negation is, intuitionistically, a notorious failure. Statements of present knowledge may be decidable, so that ${\sim}{\sim}K_t P$ would entail $K_t P$, just as intuitionists can tell what has already been proved: but statements of knowledge in the unspecified future are undecidable, just as intuitionists cannot tell what will be proved.

(14) is not evidently absurd. It forbids intuitionists to produce claimed *instances* of truths that will never be known: but why should they attempt something so foolish? For they can do justice to the platitude that not all truths will be known precisely in the formula:

(15) $\sim\forall P(P \rightarrow \exists t\, K_t P)$

For intuitionistically (15) does not entail:

(16) $\exists P{\sim}(P \rightarrow \exists t\, K_t P)$

which would compel intuitionists to produce instances ((14) and $\sim(P \rightarrow \exists t\, K_t P)$ are intuitionistically incompatible). If the second order quantification in (15) looks suspect, one can instead make claims of the form:

(17) $\sim\forall x(Fx \rightarrow \exists t\, K_t Fx)$

for suitably chosen 'F' (e.g. 'contains an even number of grains of sand').

The intuitionist status of (1) is like that of the law of excluded middle. For $\sim(P \vee \sim P)$ contradicts itself in intuitionist logic as in classical, which prompts John McDowell mistakenly to suggest that use of the former in semantics, not forcing *rejection* of excluded middle, would be compatible with realism ('Truth Conditions, Bivalence, and Verificationism', in Evans and McDowell (ed.), *Truth and Meaning*, Oxford, 1976, p. 55). But intuitionists can assert $\sim\forall P(P \vee \sim P)$, since it does not entail $\exists P{\sim}(P \vee \sim P)$. Indeed, they should do so, for any assertion of $\forall P(P \vee \sim P)$ would amount, from their point of view, to a claim of (potential) omniscience (Brouwer, 'Consciousness, Philosophy and Mathematics' in Benacerraf and Putnam (ed.), *Philosophy of Mathematics*, Blackwells, 1964): that we can see now that *we* could never be in a position to make such a claim is just what (the intuitionist explanation of the connectives says) entitles us to assert its negation. Again, if the second order quantification looks suspect, one can instead make selected claims of the form $\sim\forall x(Fx \vee \sim Fx)$: intuitionist mathematics actually has such theorems (Dummett, *Elements of Intuitionism*, Oxford, 1977, p. 84).

In short, the argument fails because it uses principles of logic that anti-realists need not – should not – accept. However, Hart does suggest a direct argument to commit intuitionists to (1). Intuitionistically, a proof of $P \rightarrow Q$ is just an evident way of turning any proof of P into a proof of Q. Take any proof of P: note underneath that you have apprehended it and so know that P: is that not just a proof that P is known? Hence $P \rightarrow KP$ would be disastrously provable for intuitionists. But that shows just that proof has not been understood in a way appropriate to intuitionism. Perhaps every proof token of P (in a mind, on paper) can be turned into a proof token that P is known: that does not entail that every proof type (as the permanent possibility of a token) can be turned into a proof type that P is known. No way leads from a structure of shareable thoughts to the further thought that some mind has contained that structure – without begging the question by assuming (1). Intuitionists say that necessarily some mind *could* contain the proof. That gives a route to (2), not (1): every proof type of P can be turned into a proof type that P is knowable. The intuitionist assertibility of (2) does show the instability of the position canvassed by McDowell: intuitionist logic with the claim 'that the truth condition of a sentence may obtain even if we cannot tell that it does, and may not obtain even if we cannot tell that it does not' (op. cit.). The intuitionist analysis of proof, without which the logic is but a dead formalism, is not innocent of anti-realist commitment. On the requisite view of proof, Hart's argument leads intuitionist proof theory into anti-realist philosophy, not into a proof of its own absurdity.

Confusingly, Dummett endorses an argument very like Hart's second (op. cit. p. 349). However, this depends on interpreting the existential quantifier so that what corresponds to our (10′) means that one has an effective means of bringing about a time t such that K_tP, not that such a time will actually come (p. 350). This interpretation is needed for the statements of the theory of the creative subject to be mathematical rather than historical in content. I have interpreted the existential quantifier in the straightforward temporal sense, since that is what Hart needs for the intuitive absurdity of (1). Thus Dummett so interprets his conclusion ((4) on p. 348) that it is analogous to our (2′) or (2), not (1). He allows his equivalent of our (14) ((4), p. 339) on either interpretation of the quantifier.

Intuitionism remains standing. That a little logic should short circuit an intensely difficult and obscure issue was perhaps too much to hope, or fear.

Trinity College, Dublin 2

© T. Williamson 1982

RUBEN ON LEWIS AND CAUSAL SUFFICIENCY

By Stig Alstrup Rasmussen

IN a recent paper, 'Lewis and the Problem of Causal Sufficiency' (Analysis 41.1, January 1981, pp. 38–41) David-Hillel Ruben contends that in his analysis of causal dependence David Lewis commits himself to the view that causal dependence involves sufficiency in the circumstances of the cause for its effect (David Lewis, 'Causation', *Journal of Philosophy* 1973, pp. 556–67; reprinted in Ernest Sosa (ed.), *Causation and Conditionals*, Oxford 1975, pp. 180–91). I.e. Lewis is supposed to hold that if e is causally dependent on c (abbreviated '$C(e,c)$' in the sequel), then not only is c necessary in the circumstances for e (abbreviated 'c nec e'), but, equally, c is sufficient in the circumstances for e (abbreviated 'c suff e'). Ruben considers, however, that a Lewis type account can in fact accommodate sufficiency only at the cost of making causal dependence a symmetric relation; at least, asymmetry will have to be ensured by some additional means. The natural proposal is that of adverting to temporal precedence; but this option is closed to Lewis, who is anxious to make room for backwards causation.

I wish to point out that Ruben's criticism is unfair to Lewis. At most, Ruben succeeds in pointing to an inconsequential fact about parsing that may, admittedly, appear to present a problem, viz. that a fairly natural reading of the conjuncts in Lewis's account of causal dependence ought not to be adopted. In consequence, Ruben affords no sound reason for disavowing Lewis's account. By the same token, there appears to be no reason for preferring Ruben's alternative proposal. Quite the contrary: as Ruben is himself aware, while Lewis's account does not, as it stands, make causal dependence a symmetric relation, Ruben's proposal does exactly that. Also, it gives rise to the difficulty of depriving that notion of a uniform content.

Lewis intends his analysis of $C(e,c)$ to capture the idea that 'whether e occurs or not depends on whether c occurs or not' (Lewis, op. cit. p. 187). This is to be effected by the use of counterfactuals, to be understood according to Lewis's *Counterfactuals* (Oxford, 1973). Lewis presents his analysis thus ('Causation', p. 187):

> The dependence consists in the truth of two counterfactuals: $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$ and $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$. There are two cases. If c and e do not actually occur, then the second counterfactual is automatically true because its antecedent and consequent are true: so e depends causally on c iff the first counterfactual holds. That is, e would have occurred if c had occurred. But if c and e are actual events, then it is the first counterfactual that is automatically true. Then e depends causally on c iff, if c had not been, e never had existed.

Thus, Lewis's analysis of causal dependence has it that:

(1) $C(e,c)$ iff $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$ and $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$.

Note that Lewis nowhere introduces the necessity and sufficiency idiom. There is no compelling reason, however, why he should oppose (2):

(2) $C(e,c)$ iff c suff e and c nec e.

Now, it is *prima facie* plausible to suppose, as Ruben evidently does, that the moment one subscribes to both (1) and (2), one is committed to (3) and (4):

(3) c suff e iff $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$.

(4) c nec e iff $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$.

Now, however, difficulties accrue. Ruben wonders what becomes of the idea of sufficiency in case c and e both occur. Nothing seems to capture this idea, since $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$ is designed to capture necessity, while the remaining counterfactual is automatically true (Ruben, p. 40). Similarly, in case neither c nor e occurs nothing seems to account for the necessity in the circumstances of the cause for its effect (ibid. p. 40, note 3).

The answer to Ruben's query is that while (2) may be adopted without detrimental effects, the same is not true of (3) and (4). Lewis's analysis does not admit of that particular parsing. On the assumption that both c and e occur, the right hand side of (3) is automatically true, but the conjunction of the right hand sides of (3) and (4) may nevertheless be taken to capture sufficiency, as well as necessity. This is because, if both c and e occur, $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$ implies the non-trivial truth of (5):[1]

(5) $\sim O(c) \mathrel{\Box\!\!\rightarrow} (O(c) \mathrel{\Box\!\!\rightarrow} O(e))$.

That is, the right hand side of (4) implies that even if c (and e) had not occurred, it would still hold that if c had occurred, then e would have occurred. Since c and e are supposed to be actual events, this does not hold automatically. It would seem, then, that

---

[1] The derivation is as follows: From

(i) $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$

and the trivial

(ii) $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(c)$,

we conclude that

(iii) $\sim O(c) \mathrel{\Box\!\!\rightarrow} (\sim O(c) \ \& \sim O(e))$.

This in turn entails

(iv) $\sim O(c) \mathrel{\Box\!\!\rightarrow} (\sim O(e) \mathrel{\Box\!\!\rightarrow} \sim O(c))$

and

(v) $\sim O(c) \mathrel{\Box\!\!\rightarrow} (\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e))$.

In view of the counterfactual equivalence of $\sim O(c)$ and $\sim O(e)$, given the counterfactual assumption that c does not occur, the consequent of (iv) may be contraposed. That is, we may conclude that

(vi) $\sim O(c) \mathrel{\Box\!\!\rightarrow} (O(c) \mathrel{\Box\!\!\rightarrow} O(e))$.

This is identical with (5). The derivation of (6) is similar. Note also that, given $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$ and $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$, both of the following are trivially true:

(vii) $O(c) \mathrel{\Box\!\!\rightarrow} (O(c) \mathrel{\Box\!\!\rightarrow} O(e))$.

(viii) $\sim O(c) \mathrel{\Box\!\!\rightarrow} (\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e))$.

(5) in conjunction with $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$ captures essential elements of the idea of causal sufficiency.[2] One might argue, therefore, that while Ruben succeeds in showing the inadmissibility of (3), he fails to demonstrate that (1) does not capture the idea of sufficiency.

Similarly, if neither c nor e occurs, then, admittedly, $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$ is automatically true. That counterfactual cannot, then, be taken to account for causal necessity; i.e. we ought to reject (4). If neither c nor e occurs, however, $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$ implies the non-automatic truth of (6):

$$(6)\ O(c) \mathrel{\Box\!\!\rightarrow} (\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)).$$

So, while (4) is inadmissible, Lewis's account of causal dependence does manage to capture the idea of causal necessity.

If one is not misled by the paraphrase encapsulated in (2) into endorsing (3) and (4), Lewis's account of causal dependence is seen to comply with the constraint that neither the necessity nor the sufficiency of causes for their effects be rendered automatically true by the mere occurrence or non-occurrence of events. Ruben's criticism makes good nothing more than the claim that, if one adopts the necessity and sufficiency idiom, one must take care to avoid subscribing to the parsing encapsulated in (3) and (4). Since Lewis nowhere commits himself to (3) and (4), Ruben's criticism is unfair. Should one incline, in view of the inadmissibility of (3) and (4), to eschew the necessity and sufficiency idiom altogether, i.e. to reject (2) as well, this would in no way upset Lewis's account of causal dependence, which is expressed exhaustively in (1). But to repeat: since Lewis may adopt (2), and since his account of causal dependence does not, as it stands, make that relation symmetric, Ruben's criticism misfires.

Now, Ruben of course holds that once Lewis endorses (2) as a preliminary rendering of causal dependence, he is committed to (3) and (4). On that assumption, Ruben is entirely right that nothing captures the idea of sufficiency, in case c and e occur. Ruben suggests that a promising way of avoiding automatic truth in this case is that of substituting (7) for (3):

$$(7)\ c \text{ suff } e \text{ iff } \sim O(e) \mathrel{\Box\!\!\rightarrow} \sim O(c).$$

Ruben concludes (p. 41) that:

> There is, then, nothing in Lewis's account that does or can be used to get at the idea of causal sufficiency$_C$ [i.e. sufficiency in the circumstances]. A lesson in this is that if an event c is necessary$_C$ and sufficient$_C$ for e, then e must be necessary$_C$ and sufficient$_C$ for c, and some alternative method will have to be found for ensuring the asymmetry of the causal relation.

---

[2] Some essential elements only: Lewis's analysis is, as pointed out by Jaegwon Kim, 'too broad to pin down causal dependency'. *Vide* 'Causes and Counterfactuals', *Journal of Philosophy*, 1973, pp. 570–2; reprinted in E. Sosa (ed.), op. cit., pp. 192–4.

Enough has been said concerning the possibility of accounting for sufficiency while avoiding symmetry of causal dependence. It deserves note, however, that Ruben's alternative proposal, if made subject to an apparently compulsory generalization, leads to the intolerable consequence of ruling out an uniform account of causal dependence. As mentioned above, Ruben's reasoning leads him to endorse (8):

(8) $O(c)$ and $O(e)$ imply that c suff e iff $\sim O(e) \mathrel{\Box\!\!\rightarrow} \sim O(c)$.

By parity of reasoning Ruben would seem compelled to adopt the following analogous principles:

(9) $O(c)$ and $O(e)$ imply that c nec e iff $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$.
(10) $\sim O(c)$ and $\sim O(e)$ imply that c suff e iff $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$.
(11) $\sim O(c)$ and $\sim O(e)$ imply that c nec e iff $O(e) \mathrel{\Box\!\!\rightarrow} O(c)$.

The point is, of course, that automatic truth of c nec e and c suff e is, in each case, avoided by adopting the appropriate counterfactual among the two pairs of *prima facie* reasonable candidates.

Now, the moment Ruben's proposal is thus generalized, the following account of causal dependence ensues:[3]

(12) $O(c)$ and $O(e)$ imply that $C(e, c)$ iff $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$ and $\sim O(e) \mathrel{\Box\!\!\rightarrow} \sim O(c)$;

and

$\sim O(c)$ and $\sim O(e)$ imply that $C(e, c)$ iff $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$ and $O(e) \mathrel{\Box\!\!\rightarrow} O(c)$.

Since counterfactuals do not admit of contraposition, (12) fails to provide an unitary account of causal dependence. Indeed, this perfectly natural generalization of Ruben's proposal renders the analysis of that relation contingent on what happens to be the case, i.e. on whether or not c and e occur.[4]

University of Copenhagen
Købmagergade 50
DK-1150 Copenhagen K, Denmark

© STIG ALSTRUP RASMUSSEN 1982

---

[3] In fact (12) is only a partial definition of causal dependence, since it does not cover all cases. On Lewis's account, $C(e, c)$ is automatically false, in case $O(c)$ and $O(e)$ differ in truth value. On Ruben's proposal, however, this must be ensured by expressly adding a clause to that effect. One might, for instance, add either of the following clauses:

(i) If $O(c)$ and $O(e)$ differ in truth value, then $C(e, c)$ iff $\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$ and $\sim O(e) \mathrel{\Box\!\!\rightarrow} \sim O(c)$.

(ii) If $O(c)$ and $O(e)$ differ in truth value, then $C(e, c)$ iff $O(c) \mathrel{\Box\!\!\rightarrow} O(e)$ and $O(e) \mathrel{\Box\!\!\rightarrow} O(c)$.

It is easy to check that, on either alternative, $C(e, c)$ is guaranteed to be false in case $O(c)$ and $O(e)$ differ in truth value. So, (12) may be supplemented by either.

[4] I am indebted to the Editor for helpful comments on an earlier version of this paper.

KNOWING, BELIEVING, AND GUESSING

By ROY A. SORENSEN

IN 'Knowledge – by examples' (ANALYSIS 27.1 October 1969, pp. 1-11), Colin Radford argues against the entailment thesis: if one knows that p then one believes that p. Radford gives the example of Jean, a French-Canadian who believes that he knows nothing about English history. To humour a friend, he agrees to answer some questions about English history. All along, he insists that he is merely guessing and so does not believe any of his answers. Yet, Jean does very well. Radford invites us to conclude that Jean could not have been merely guessing and must have known the answers despite the fact that Jean did not believe them. Radford makes the story more plausible by adding that Jean then remembers that he did study English history years ago.

I think that Radford's example owes whatever persuasiveness it has to the plausibility of the following principle:

(A) If one merely guesses that p, then one can only be accidentally right that p.

The plausibility of (A) is undermined by another example involving an unconfident examinee. Julia has fallen far behind in her physics class and so her friend helps by teaching her a method of calculating viscosity. Her friend warns her that the method is only reliable up to the second decimal place. The next day Julia is given a test consisting of ten problems requiring viscosity calculation to the third decimal place. Since Julia is only familiar with the method taught to her by her friend, she decides to use it even though she believes that the probability of it providing the correct answer on a given question is only .1. Her teacher is astonished when he learns that Julia answered each of the questions correctly. When the teacher asks Julia how she did so well, Julia honestly replies that she guessed with the help of her friend's method. After Julia describes the method, her teacher informs her that, contrary to her friend's warning, the method is reliable up to the third decimal place. Even though Julia used a reliable method, she did not know the answers to the examination since she was mistaken about the reliability of the method.

This example shows how one can be non-accidentally right by merely guessing, and so undermines (A). In the case of Jean, the method is the informal one of adopting the first answer that comes to mind. If one has learned English history, this is a reliable method. But if one knows nothing about English history, this is an unreliable method. Since Jean falsely believes that his method is unreliable, Jean does not know the correct answer to the English history questions. Thus, the case of Jean is not a counterexample to the entailment thesis.

Michigan State University,
East Lansing, Michigan 48824, U.S.A.

©ROY A. SORENSEN 1982

GOING NOWHERE FAST?

By ROBERT ELLIOT

CHRISTOPHER RAY (ANALYSIS 42.1, January 1982) claims I am not the person I think I am. Andrew Brennan (ANALYSIS 41.1, January 1982) lends him some support but in the end suggests that there may be no fact of the matter. There are several things I would like to say in my defence (see my 'How to travel faster than light', ANALYSIS 41.1, January 1981).

I

Both Ray and Brennan correctly observe that my faster than light journey has left me with no memories of the time spent on Alpha Centauri. Indeed the time I spent there has had no impact on my consciousness whatsoever since the causal connections which might have made this possible are, *ex hypothesi*, non-existent. There may be some temptation to think that this destroys psychological continuity between the person who departed Alpha Centauri and the person who arrived on Earth in literally no time at all. Moreover, so the suggestion goes, without psychological continuity there can be no question of my having survived the journey. However, there is psychological continuity of a non-standard, thought not unfamiliar, form.

I presume that Ray and Brennan are aware of the operations of Fantasy Enterprises and of the difficulties that beset the company in its early days. Perhaps a reminder is in order. Fantasy Enterprises, as most of us know, runs trips to the more outlandish planets where anything and everything goes. Quite a few members of the first expeditions suffered profound psychological damage as a result of their unusual experiences. To avoid such continuing damage, and the associated law suits, Fantasy Enterprises introduced the erasure technique. All members of their expeditions must agree to have all memories and all other psychological consequences of their visits to perverse places erased prior to their return to Earth. This is accomplished by adjustments carried out at the extra-terrestrial transmission stations.

There can be little doubt that Fantasy Enterprises brings its clients back alive. The case is in principle no different from one in which the erasure technique is used on a non-traveller in order to obliterate the memory of some traumatic experience. What goes on can be represented more formally by taking three time-slices of the one person, P1, P2 and P3. P1 is standardly, psychologically continuous with P3 and with P2. P2 and P3 are not standardly, psychologically continuous. However, since P2 and P3 are both standardly, psychologically continuous with P1 we can regard them as indirectly, psychologically continuous with each other. That this form of psychological continuity is an adequate support for personal survival (indeed identity) is evidenced by a consideration of the erasure technique as it is used in the cases referred to just now. The cases involve what is essentially just an exaggerated lapse of memory.

Now the only difference between the Fantasy Enterprises expeditions and my trip is that the former consistently employ the beam-transport method. This aside the cases are structurally the same, the relationshhips between corresponding time-slices of persons are isomorphic and so on. Ray and Brennan cannot merely appeal to the fact that my journey was faster than light in order to show how it is that Fantasy Enterprises brings its clients back alive whereas 'I' didn't quite make it. This simply begs the question. Perhaps what misled Ray and Brennan is that the erasure technique seems unavoidably built into the faster-than-light mode of travel. Even if it were logically built in I cannot see that it would alter matters. Indeed it might provide a secure guarantee that one's terrestrial life is minimally interfered with by inter-stellar travel. Not only does faster than light travel minimally interfere with terrestrial friendships, marriages, and so on because it is the fastest mode of travel, it minimally interferes with them because it reduces to zero those psychological changes which bring about estrangement from friends, spouses and so on.

It is worth pointing out that the story which inspired my faster than light travels, namely Xenon's story, is not open to Ray's and Brennan's (misplaced) criticism in the way that my case is. Xenon travelled faster than light and incurred no psychological costs. Thus erasure is not logically built into faster than light travel. It is not even logically built into non-fortuitous faster than light travel as I shall now demonstrate.

II

The erasure problem, if it is a problem, is a function of the stage of technological advance. In fact there are presently breakthroughs being made by researchers at the La Place Institute of Computer Science which promise to put paid to Ray's and Brennan's complaints.

The matter is far too complex to be dealt with in detail. However, it will soon be technically possible to predict with extreme accuracy the detailed psychological impacts that trips to other star-systems will have on any given traveller. The relevant programme requires vast amounts of data and the programme itself is enormously complex but the predictions are feasible and will soon be commonplace. Such predictions will provide the perfect antidote to built in erasure. Ray and Brennan fail to distinguish between the familiar and the possible. It is the latter to which we should look if we wish to map the logical boundaries of necessary and sufficient conditions for personal survival.

III

I noted at the outset that Brennan has a different reaction to faster than light travel from Ray's despite his use of an argument similar to Ray's. In fact Brennan has drawn something like the intended moral from the tale.

It is difficult to give up the belief that persons are determinate, that identity and survival are all or nothing matters and that there is a self which persists through time. My intention was to raise doubts about these matters and to suggest that once one abandons a bodily continuity criterion for personal survival in favour of a plausible mentalistic criterion, one can be brought to a point where these beliefs must be seriously doubted. I do not want to suggest that this counts in favour of the bodily continuity criterion. I do not think that it does. What I primarily intended was to highlight tensions within our conceptual scheme and its inadequacy in representing the world. Fantasy as well as fact provides a means of doing this, sometimes a better means. It is appropriate to conclude with a remark of Thomas Nagel's: '... it is possible that the ordinary, simple idea of a single person will come to seem quaint one day, when the complexities of the human control system become clearer and we become less certain that there is anything very important that we are *one* of. But it is also possible that we shall be unable to abandon the idea no matter what we discover.' (Thomas Nagel, 'Brain Bisection and the Unity of Consciousness' in his *Mortal Questions*, Cambridge University Press 1979, p. 164).

Brisbane College of Advanced Education
Kelvin Grove, Brisbane, Queensland 4059
Australia

© ROBERT ELLIOT 1982

ACTIONS, REASONS AND CAUSAL EXPLANATION

By DENISE MEYERSON TAYLOR

YOU can causally explain bridges collapsing by talk of hurricanes, as you cannot causally explain the occurrence of the event reported on p. 13 of the *Tribune* by saying it was caused by the event reported on p. 5 of *The Times*, even if it was caused by the event reported on p. 5 of *The Times*.[1] Peter Smith, in his paper 'Hess on Reasons and Causes' (ANALYSIS 41.4, October 1981), claims that when actions are rationalized by citing beliefs and desires, that is like explaining collapsed bridges by mention of hurricanes. This in two respects: like hurricanes, beliefs and desires are, first, 'causally salient', but second, not fit to appear, described thus, in tough-egg causal laws. But those two things cannot, or so I shall argue, both be true of beliefs and desires.

Causation is a relation between events. But causal salience is, I take it, a relation between statements, a relation between events as described in certain ways, i.e. it is causal explanatoriness. But I find no argument in Smith for the claim that beliefs and desires, so described, do causally explain actions. That they don't is indeed the claim of the best reading of the argument that beliefs and desires are 'logically connected' to actions. The bad reading of that argument is obviously flawed. Necessarily, the cause of *b* caused *b*; but it doesn't follow that the cause of *b* is necessarily connected to *b*. Similarly, even if it is the case that, necessarily, intentions lead to actions (with the usual provisos about opportunity, information, constancy of desire, etc.), it doesn't follow that the events which are intentions are necessarily connected to actions, that is, that they don't cause actions. But the more plausible version of that argument is, though wrong, not as obviously wrong. It is that, necessarily, intentions lead to actions; therefore citing intentions cannot causally explain actions (though citing neurophysiological events might causally explain contractions of muscles). Or, to put this in Smith's terminology, intentions, while causes of actions, are not, under their mental descriptions, causally salient to actions. Citing an intention, in other words, is, on this better reading of the logical connection argument, as little use as attempting to explain dissolvings by citing solubility.

I think that citing intentions can causally explain actions, as citing water-solubility cannot causally explain dissolvings in water, partly because it isn't part of the meaning of 'intention' that, other things being equal, intending to do something is followed by doing it. There is evidence for intentions which is independent of their alleged manifestations (see D. Davidson, 'Actions, Reasons, and

---

[1] I am grateful to David Brooks, Mark Leon, Alan Ryan and a referee of *Analysis* for useful comments, and to Paul Taylor for his generous help.

Causes', in *Philosophy of Action*, ed. A. R. White (Oxford, 1968), p. 90) so that there isn't here a bogus explanation of the *virtus dormitiva* sort. But not only does Smith not say why *he* believes that citing intentions can causally explain actions; it is also the case that he oughtn't to believe it. Briefly: I shall argue that beliefs and desires, so described, can causally explain actions, but that is because they can figure under those descriptions in causal (though not mechanistic) laws. But Smith thinks they *cannot* so figure. That thesis (the anomaly thesis) in conjunction with two other of Davidson's theses, the token-identity thesis and the-impossibility-of-psycho-physical-laws thesis, entail that beliefs and desires cannot, so described, causally explain actions, though they will, of course, under those descriptions, justify them.

Why I say that is this. As Davidson points out, '*A* caused *B*' entails only that there be some causal law instantiated by true descriptions of *A* and *B* and not that there be a law relating *A* and *B* under the descriptions '*A*' and '*B*' (ibid., p. 92). But what is entailed by '*A* causally *explains B*'? Knowledge of a law isn't necessary for causal explanation and the existence of a law, while necessary, isn't sufficient for causal explanation. How can there be explanation when there isn't knowledge of the relevant law? Davidson shows that it is an acceptable explanation of a window's shattering that a rock was flung at it, although the generalization, 'glass is fragile and tends to shatter when struck quite hard', is not law-like, and the law which supports it will be couched in quite other terms. But nor is mere existence of a law sufficient for explanation. For '*A* caused *B*' to explain why *B* happened, the cause and effect must be cited in terms which lay bare a connection with at least a loose generalization. That is why it is no explanation of the bridge's collapsing that it was caused by the event reported on p. 5, when that event is described like that, even though there exists a law relating those events under some other descriptions. For the citing of *A*, thus described, causally to explain why *B*, thus described, occurred, there must at least be acquaintance with a loose generalization relating *A* and *B* under the descriptions '*A*' and '*B*'.

But from the fact that some (known) true accidental generalizations make the adducing of a cause explanatory, it doesn't follow that all do. *Is* knowledge of a true accidental generalization (relevant to the singular causal statement) sufficient for explanation? I want to argue that it isn't but that it is is just what Smith seems to assume. There are generalizations which make use of the concepts of reasons and actions whereas there aren't generalizations about events classified according to where you will read about them in the newspaper. The psychological generalizations are, of course, accidental generalizations for Smith (the anomaly thesis), but he can only think that that doesn't matter to their grounding the explanatoriness of singular causal statements on the analogy of something like the fragile glass generalization. But some true, accidental generalizations are not explanatory. It is no explanation of why a chair in my room is broken that all the chairs in my room are broken. Of course, that is for the trivial reason that there isn't an underlying law here at all. But another way in which a true accidental generalization can fail to facilitate explanatoriness is where, or so I shall suggest, though there is a law, there isn't the right sort of relationship between the terms used in the generalization and the terms used in the law; there isn't a systematic connection between the law and the accidental generalization, because there exists only a token-identity between the events falling under the generalization and the events falling under the law. Smith gives us no reason to think his accidental generalizations aren't like that, and indeed, for him, they must be.

What I suggest is this: for the citing of *A*, thus described, causally to explain why *B*, thus described, occurred, it is necessary that, either

(1) the supporting law relate *A* and *B* straightforwardly under the descriptions '*A*' and '*B*',

or,

(2) if the generalization in which the descriptions '*A*' and '*B*' figure is accidental, then the type-identity statements, '*A* is *C*' and '*B* is *D*' be true, where '*C*' and '*D*' are the descriptions of *A* and *B* used in the law.

Consider the hurricane case. Smith is right to say that although there exists no law relating the occurrence of hurricanes to the collapsing of bridges, still it is in order to explain the bridge's collapsing by talk of the hurricane. But that is because there is a systematic way of translating from the vocabulary used in the everyday generalization into the vocabulary used in the law. A hurricane is a type of physical phenomenon and the law which does exist specifies in physicalist terms what type that is. The physical specification gives the nature of the thing. That is why it is explanatory to cite the hurricane: we do so in a *faute de mieux* way, in the knowledge that the mapping from everyday talk to theory exists at least in principle. That doesn't prevent the generalization which makes use of the pre-scientific concepts from being accidental. The scientific descriptions give the nature of the things referred to by the pre-scientific descriptions, but the latter descriptions are themselves too vague for the events they describe thus described to instantiate a law. All we know is that a hurricane of around a certain force will destroy a bridge of around a certain size, and there is no way of tightening that up in that vocabulary so as to make it exceptionless. The general point is that laws do not deal with events just as such, but with events as described in certain ways. As Davidson points out, if there is a law that connects *a* and *b* as such, and if $a = c$, it doesn't follow that there is a law that connects *c* and *b* as such (D. Davidson, 'Reply to Attfield', *Essays on Actions and Events*, p. 242, Oxford 1980). It follows from all this that the hurricane case satisfies the conditions for option (2).

But now consider beliefs and desires. The token-identity theory goes along with the view that there is no particular physical state which has to realize a given mental state, although a mental state must be realized by some physical state or other. The autonomy thesis says that what psycho-physical generalizations there are are not law-like. It follows from these two claims that statements relating beliefs and desires to actions cannot be systematically related to statements expressed in a physicalistic vocabulary, for it is a contingent matter as to how a kind of mental state is realized physically. That is, the connection between beliefs and desires and physiological states is purely accidental. So option (2) isn't open: there is no explanation going on here but only coincidence. But according to Smith, the laws relating beliefs and desires won't deal in *those* concepts (the anomaly thesis again). So option (1) isn't open either. I conclude that if the anomaly thesis is right (and if Davidson is right about token-identity and autonomy as well), then beliefs and desires cannot causally explain actions.

I agree that there are no laws relating neurophysiological and mental events (that is, I agree that option (2) isn't open here), but I believe that there *are* causal laws couched in psychological terms, though these won't be mechanistic because, to use Bennett's terminology, their antecedents will use instrumental rather than intrinsic predicates. Briefly: any R thing will do whatever it can which it registers as required for and sufficient for its becoming or remaining G. The antecedent condition, that a particular bit of behaviour, B, is (registered as) necessary and sufficient for G, explains the occurrence of B (J. Bennett, *Linguistic Behaviour*, Cambridge, 1976). Mechanistic laws of course make no reference in this way to a later time, but it would appear dogmatic to say that only they adduce a cause. If this is true, citing beliefs and desires under their mental descriptions meets the conditions for option (1), and *that* is why talk of them causally explains action.

University of Cape Town
Rondebosch 7700, S. Africa

© DENISE MEYERSON TAYLOR 1982

BAD NEWS FOR ANOMALOUS MONISM?

By PETER SMITH

IN his paper 'The Argument for Anomalous Monism' (ANALYSIS 42.1, January 1982), Ted Honderich argues that the Davidsonian position has grave difficulty in accommodating a central conviction which any acceptable philosophy of mind should surely underwrite, namely our conviction that the mental is causally efficacious.

Davidson's type of monism is intended to reconcile the following propositions which are prima facie inconsistent. (1) There are causal connections between mental events and physical events. (2) Wherever there is a causal connection, there is a covering law. And (3), there are no psycho-physical laws. As is now well known, Davidson aims to reconcile these three propositions by noting that we must interpret the second one in a weak sense – i.e. as asserting only that, where there is a causal connection between events, there must be some descriptions of these events in virtue of which they instantiate a law. A given singular causal claim need not identify the events in question by means of concepts which are fit to feature in the supporting general law. This opens up the possibility that it could be true that Jane's desire caused her action consistently with there being no psycho-physical laws, by virtue of the particular desire and the action falling under physical descriptions which *do* instantiate a law. This possibility presupposes the identity of the mental with the physical: and the monism here (if we are to avoid the re-introduction of psycho-physical laws) must be anomalous.

Honderich argues that this familiar complex of ideas is not able to secure the causal efficacy of the mental. For the point that a singular causal claim need not be couched in terms which recur in the backing law is just the point that we can identify a cause and effect by means of properties which are not relevant to their causal linkage. Davidson is thus committed to distinguishing, among the properties of a given event, those in virtue of which it causes a certain upshot – and these properties will be the ones governed by the relevant background causal law. In the light of this, let's ask which properties of Jane's desire are relevant to its causing her action. The immediately plausible reply is that it is in virtue of its mental properties that the desire causes the action: or in Honderich's phrase, it is 'the mental as mental' which causes the physical. But this reply implies that there are, after all, nomological ties between mental properties and physical ones, and we are back with the unwanted psycho-physical laws. So suppose instead that we say that it is 'the mental as physical' which causes the physical: in other words, suppose we maintain that the properties of Jane's desire which are relevant to its causing the action in question are among its *physical* properties. Note, however, that if we are to continue to avoid the re-introduction of psycho-physical laws, then the connection between the mental and physical properties of a given event must be deemed *not* to be nomological. And according to Honderich 'that is to say that the connection between the mental and the physical is accidental'. So it follows that, in particular, the connection between an event's having the property of being a desire and its causally relevant physical properties is accidental. But this conclusion seems intolerable; as Honderich puts it, 'Here, it seems, we do not get the efficacy of the mental'.

All this, claims Honderich, is bad news for anomalous monism. But it is worth pointing out that, whatever the merits of the argument, it certainly is not news. For here is Fred Stoutland writing in 1976:

> ... On Davidson's theory, the causal powers of propositional attitudes have no determinate connection with the fact that they *are* propositional attitudes. It is, on the one hand, because they are propositional attitudes that they account for intentionality (or figure in the explanation of intentional action); it is, on the other hand, (only) because they are identical with physical events that they cause behaviour. But that any particular event should be both a particular attitude and a particular physical event – that is, that the same event should have both a propositional attitude description and a physical description – is simply a brute fact for which there cannot, on Davidson's grounds, be an explanation. ... [To avoid re-introducing psycho-physical laws] it must be a brute fact, without explanation, that an attitude causes the behaviour it does.[1]

The argument here is evidently the same as Honderich's; and Stoutland too, reasonably enough, finds its conclusion unacceptable. But does Davidson's anomalous monism really commit him to this conclusion? I think not.

The crucial move in the Stoutland/Honderich argument is their inference from the lack of any nomological connection between the mental and the physical to the claim that it is an 'accident' or 'brute fact' that a given event has the particular combination of mental and physical properties it displays. And this move should be resisted.

Let's begin by asking the anomalous monist how we are to pick out the physical event (or state) which is identical with a given mental event (or state). In other words, given a mental specification of an event, how are we to determine which physical event satisfies the specification? Of course, we cannot demand from the identity theorist a recipe which can in practice be applied in order to identify empirically the physical realization of a given mental state. The demand is the weaker one that we are provided with some non-arbitrary and well-motivated completion for such a schema as 'The physical state which is X's belief that p at time t is that physical state such that . . .', where the filling does in principle pick out a unique state.[2]

---

[1] Frederick Stoutland, 'The Causation of Behaviour', in *Essays on Wittgenstein in Honour of G. H. von Wright* (ed. J. Hintikka), *Acta Philosophica Fennica*, vol. 28, 1976, pp. 307–8. Stoutland expounded the argument in a paper read in Oxford and elsewhere in 1975, a descendant of which appears as 'Oblique Causation and Reasons for Action', *Synthese*, vol. 43, 1980.

It seems clear enough how the monist must proceed here. He can hardly offer to identify *X*'s belief with some physical state which lacks the right sort of causal connections with the behaviour we interpret as done because of that belief; for this would ratify what are, by ordinary standards, quite the wrong claims of the form '*X* did *A* because he believed that *p*'. So, putting it crudely, we must identify *X*'s mental states with physical states which are causally appropriate. Refining this rough idea, we arrive at the following suggestion: the physical state which is *X*'s belief that *p* (for example) is that state which is (a) causally dependent on those antecedents which folk psychology recognizes as explaining *X*'s belief (such as perceptual input), and such that (b) it is causally involved in the production of such behaviour as folk psychology interprets as actions done because *X* believes that *p*.

Now it is very easily seen that, if the anomalous monist does identify the physical state which is *X*'s belief in this sort of way, then he is not vulnerable to the Stoutland/Honderich argument. Suppose that *X* flipped the switch because he believed it would turn on the light. Then the physical state which is *X*'s belief is that state (whatever it is) which has the right causal relations to the antecedents and consequences of *X*'s belief: and so in particular, *X*'s belief is that physical state which (among other things) is causally involved in the production of the behaviour which we interpreted as *X*'s flipping the switch because he believed it would turn on the light. So – very trivially – the physical state which is *X*'s belief caused *X*'s action. It is not an 'accident' or 'brute fact' that the belief (in virtue of its physical properties) caused the action; for the physical state which is the belief is partially *identified* as the state which has the right physical properties to cause the action! And note that none of this reinstates unacceptable psycho-physical laws. There is no route, from the thought that if *X* on a given occasion believes that *p* then he is in that particular physical state which has the antecedents and consequences of his belief, to the conclusion that there must be some general law-like correlation between *X*'s believing *p* and a type of physical state directly identified in purely physical terms.

In short then, it can both be the case that there are no nomological links between the mental and the physical and yet also the case that it is no accident that *X*'s belief (in virtue of its physical properties) causes the action it rationally explains. So the Stoutland/Honderich argument entirely collapses.

Against all this, it might be protested that the anomalous monist's method of identifying the physical state which is, for example, *X*'s belief that *p* is in the present context something of a cheat. For the operative clauses (a) and (b) above help themselves to the 'because' of folk-psychological explanation, and it was worries about the role of causality here that started off the whole argument in the first place. It is important to see how this implied accusation of circularity can be rebutted.

---

[2] For more on the legitimacy of this demand, see my paper 'On Identifying the Mental and the Physical' (*Canadian Journal of Philosophy*, forthcoming).

The accusation is presumably grounded in something like the following thought. We have tried to rescue the anomalous monist from the Stoutland/Honderich argument by suggesting that the physical state which is *X*'s belief (and hence the causally relevant physical properties of his belief) is fixed in part by the requirement that it causes actions done because of that belief. But fixing which actions are done because of a given belief presupposes that we can establish some causal claims about beliefs: and – in the hypothesized absence of psycho-physical laws – this means bringing to bear some purely physical laws. However, these physical laws can only be applied if we have *already* set up an identification of a given belief with a specific physical state. So, on pain of circularity, this identification cannot in turn presuppose that we have already fixed which actions are done because of a given belief.

The obvious first reply to this line of thought is that it just isn't generally true that establishing causal claims requires 'bringing laws to bear'. I can know that on this occasion dropping the plate caused it to smash, without benefit of access to the relevant laws concerning e.g. the effect of stress on crystalline structures. Some general rules of thumb, falling well short of hard-edged laws, suffice to bolster my everyday bits of causal knowledge.[3] Likewise, folk-psychology provides rough generalizations which are not themselves law-like but which (it could again be argued) suffice to provide grounds for particular causal claims about beliefs and desires. This enables us to break into the supposed circle.

This first line of reply is sufficient as it stands. However, it is perhaps worth noting in conclusion the possibility of a more ambitious attack on the accusation of circularity. Let's consider more carefully the sense in which folk-psychology makes causal claims. Two views need to be distinguished here. The stronger one is that a claim such as '*X* flipped the switch because he believed it would turn on the light' provides a causal *explanation* of *X*'s action. The more modest view is that such a claim mentions causes but is not itself a causal explanation. More generally, while beliefs are indeed states which cause actions, they do not rationally explain the action in virtue of being such states: rather it is the other way about – i.e. the belief is that physical state which causes the actions it rationalizes in virtue of being physically identified in precisely such a way as to fit that function. If the anomalous monist adopts this second view, then all appearance of circularity in his position evaporates.

[3] See Davidson in *Essays on Actions and Events*, e.g. at p. 224.

This is not the place to rehearse in detail the virtues of the more modest view, or to discuss its relations to Davidson's position. Sceptics about the coherence of such a view might, however, reflect on the related case of teleological explanations in the light of Jonathan Bennett's fine discussion in his *Linguistic Behaviour*. There Bennett argues that a claim such as 'The missile swerved to the left because it registered that its target had swerved' could provide a genuine explanation of the missile's trajectory, and there is a good sense in which this explanation is of a different type from causal mechanistic explanation. But, of course, the event which is the missile's registering of its target's change of course is doubtless a physical event. Suppose, then, we ask 'which physical event?' The evident answer is: the one which is responsible for, among other things, making the missile swerve. So, the registering causes the swerve. In short then, teleological explanations are not causal explanations although they mention causes. The anomalous monist might well commend this as a model for our understanding of folk-psychological explanations, thus preserving the causal efficacy of the mental while side-stepping Honderich's attack.

University College of Wales
Aberystwyth, Dyfed SY23 3DY

© PETER SMITH 1982

REPLY TO FORBES

By WILLIAM SEAGER *and* GRAEME HUNTER

I

EVERYONE has recognized inept sub-titling of foreign films, and this without recourse to a manual of canonical translation. Graeme Forbes (in 'Canonical Counterpart Theory', this journal 42.1, January 1982) argues that the situation is somehow different with regard to the languages of modality. Specifically, he rejects the view set forth in 'The Discreet Charm of Counterpart Theory' (this journal, 41.2, March 1981), in which we argued that even in these cases such manuals are unnecessary.

Forbes believes instead that 'sentences of Lc [the language of Counterpart Theory (CT)] are intended to interpret or elucidate the meanings of Lm sentences [first-order modal language]' (op. cit. p. 33). To this end he attempts to provide an adequate formal translation scheme from Lm to Lc. We think it to be no more plausible that Lc is *intended* to interpret Lm than that English is intended to interpret French. Lewis himself, who begat CT (in 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy*, 65, 1968) gives no indication that this was his intention, characterizing the two systems as '*rival* ways of formalizing our modal discourse' (op. cit. p. 116, our emphasis). In our view the motivation of CT is to provide an apparatus for expressing modal facts. One of the striking advantages of CT is its ability to express these facts within a marginally augmented but otherwise standard first-order predicate logic syntax, for which it provides (implicitly) a new and intuitive semantics. This is not to deny that a translation scheme is possible from Lm to Lc (though the reverse is clearly impossible); we intend only to deny its necessity. Even to argue that a given scheme is correct is to appeal to criteria not given by the scheme, but presupposed by it, thus making the scheme logically unnecessary, however desirable it may be for practical or systematic purposes.

II

There may well still be an acceptable translation scheme, but Forbes's, like Lewis's, appears to have distressing consequences. We will note a particularly disturbing one.

Allen Hazen objected to Lewis's translation scheme (in 'Counterpart-Theoretic Semantics for Modal Logic', *Journal of Philosophy*, vol. 76, pp. 319-38) because it failed to validate the inference from:

(1) LRab

to:

(2) L(Ex)Rax.

Our Counterpart-Theoretic solution to this problem was to note the multiple ambiguity of the Lm-sentence 'LRab' – only on certain readings of this would the desired consequence logically follow. We gave these Counterpart-Theoretic readings. The recognition and resolution of ambiguity seemed to us an advantage of CT. Now Forbes claims that our 'four readings correspond to four different sentences of Lm ...' (op. cit. p. 36). This is not clear to us. On Forbes's scheme any sentence which contains terms a and b within a single necessity operator will have the conjunctive antecedent 'Cxaw & Cybw' ('C' stands for '... is a counterpart of ... at world ...'). Some of our readings have antecedents with but one of these conjuncts. Thus we find it difficult to see how Forbes, ingenious though he is, can produce four Lm sentences with the appropriate semantical qualities.

In any event, there is clearly a sense of (1) from which (2) logically follows. Forbes admits that his canonical translation of

(1), viz. (using Forbes's numbering):

(13) (w) (z) (t) Czaw & Ctbw → Rztw)

does *not* entail his translation of (2):

(14) (w) (z) (Czaw → (Ex) (Exists(x, w) & Rzxw)).

Forbes notes that the inference can be validated only with the addition of the so-called Falsehood Convention (FC), explicated as, 'all non-logical atomic sentences are false at worlds where the things named in them do not exist' (op. cit. p. 36). This is so because on Forbes's scheme any given thing, a, has a counterpart, c, *at* every world, w, where the 'at' signifies either that c is in w (i.e. Exists (c, w) in Forbes's notation) or c = a if a has no counterparts in the world at issue. Thus if FC was not in force (13) could be true while (14) was false when the 'Rzxw' of (14) holds but x is not *in* w. Thus, in order to validate the inference from (1) to (2) in CT, and to answer Hazen's objection, Forbes must impose FC.

We note, however, that this has the disastrous consequence that nothing can have any essential property without being a necessary existent. This is easy to see, for FC implies that if in any world, w, a does not have a counterpart, the sentence 'LFa' is false.

Formally, the Forbes's canonical translation of LFa would be:

(3) (w) (x) (Cxaw → Fxw).

Suppose now for reductio that —L(Ex) (x = a). This, on Forbes's scheme, translates as:

(4) —(w) (x) (Cxaw → (Ey) (Exists(y, w) & y = x)).

(4) trivially entails:

(4′) (Ew) (Ex) (Cxaw & —(Ey) (Exists (y, w) & y = x)).

It is then easy to derive a contradiction from (3), (4′) and the relevant Counterpart Theoretic formalization of FC:

(FC) (x) (w) (Fxw → Exists(x, w)).

Such a result is completely unacceptable. The mere assumption that Socrates is essentially a human being should not imply that Socrates exists necessarily. Our approach does not have this difficulty; it also involves no mechanical translation scheme. We let the reader draw the moral.

University of Toronto
Toronto, Canada M5S 1A1

© William Seager *and* Graeme Hunter 1982

MIND

A QUARTERLY REVIEW OF PHILOSOPHY
edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

Vol XCI No 364 October 1982

Articles
The Trouble with Level-Generation: H. McCANN
How Time Flies: G. SCHLESINGER
Causation, Sensations, and Knowledge: W. ROBINSON
Meaning, Use and Privacy: E. CRAIG

Discussions
The Euthyphro Dilemma: M. MacBEATH
Utility and Distribution: M. McDERMOTT
On the Alleged Necessity of True Identity Statements: E. LOWE
Hume's Argument for the Dependent Existence of Perceptions:
A. PAVKOVIĆ an Alternative Reading:
Mackie on Personal Identity: G. DOORE
A Note on Frege's Notion of Wirklichkeit: A. REIN

Critical Notice
L. Wittgenstein, *Remarks on the Philosophy of Psychology, Vols. I and II:* H. MOUNCE

Book Reviews

Books Received

Orders to:
Journals Department,
Basil Blackwell Publisher Ltd.,
108 Cowley Road,
Oxford OX4 1JF, England

Annual subscription:
Institutions: £10.00 or $22.50 (Canada $26.00)
Individuals: £8.00 or $18.00 (Canada $21.00)
Single issues: £3.00 or $6.75 (Canada $7.75)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

JOURNAL FOR THE THEORY OF SOCIAL BEHAVIOUR

General Editor: P. F. SECORD Co-Editor: P. COLLETT
Associate Editors: A. GIDDENS, P.MANICAS, C. SMITH
Managing Editor: J. S. WEST

The aim of the journal is to bring together in fruitful debate all social scientists and philosophers who might contribute to the development of theory and method in the study of individuals in society.

Recent contributions include: Isaac Franck on psychology as science; Martin Bunzl on the meaning of 'meaningful behaviour'; John L. Smith a structuralist interpretation of the Fishbeinian Model of Intention; E. Kayser and T. Schwinger on the justice concept.
Published March, July and October.

Annual subscriptions: Institutions £23.00 (inland) £28.25 (overseas) $59.00 (USA) $70.75 (Canada); Individuals £12.00 (inland) £14.75 (overseas) $34.00 (USA) $39.00 (Canada)

Orders to: JOURNALS DEPARTMENT, **BASIL BLACKWELL PUBLISHER**.,
108 Cowley Road, Oxford OX4 1JF, England.

The Philosophical Quarterly

Vol. 32 *No.* 129 *October* 1982

The Existence of Social Entities — *David-Hillel Ruben*
Aristotle on Inefficient Causes — *Julia Annas*
Conversation and Conditionals — *Anthony Appiah*
God's Law and Morality — *Stephen R. L. Clark*

DISCUSSIONS
There are Absolute Rights — *Alan Gewirth*
Intentionality: A Reply to Stiffler — *E. J. Lowe*
A Mistaken Argument — *Chris Mortensen*
Opportunity Rights? — *John Robertson*
Opportunity — Right! — *Michael Levin*

CRITICAL STUDIES
Hacker and Baker on Wittgenstein (P. Hacker & G. Baker, *Wittgenstein: Understanding & Meaning*) — *David Bell*
Pure Morality and Impure Truth (N. Cooper, *The Diversity of Moral Thinking*) — *Jonathan Harrison*

BOOK REVIEWS — BOOKS RECEIVED

The Editors intend to devote one issue in **1983** to articles on **Medical Ethics** and one in **1984** to articles on **Frege.** Contributions should be submitted, **by 31 December of the preceding year,** to The Executive Editor, The Philosophical Quarterly, The University of St. Andrews, Scotland.

Annual Subscriptions:
Individuals £8.00 ($20.00) Institutions £16.00 ($40.00)
Single issues: £4.00 ($10.00)
for orders placed directly with the Journals Manager, Scottish Academic Press, Ltd., 33 Montgomery Street, EDINBURGH EH7 5JX, Scotland.

Philosophical Books

***Philosophical Books** aims to provide prompt, scholarly reviews of professional books in philosophy. Besides conventional reviews in both contemporary philosophy and the history of philosophy, it now includes regular Discussion features, where the authors of selected titles have the opportunity to reply to reviewers. There are occasional subject surveys and issues devoted to a single theme.*

Books for review should be sent to the Editor, Dr M.A. Stewart, Department of Philosophy, University of Lancaster, Lancaster LA1 4YT.

Subscription prices are as follows:

Institutions:
£16.50 (UK) £20.75 (overseas) $41.00 (US) $49.25 (Canada)
Individuals:
£10.50 (UK) £13.00 (overseas) $26.00 (US) $31.25 (Canada)
Single issues:
£4.50 (UK) £5.40 (overseas) $11.25 (US) $13.50 (Canada)

*There are now **four** issues a year.*
Single copies and back issues are available.

BASIL BLACKWELL PUBLISHER
Journals Department. 108 Cowley Road, Oxford OX4 1JF

RATIO

edited by
MARTIN HOLLIS

Ratio takes a resolute stand against scepticism and irrationalism and pursues the perennial goal of rational philosophy. It deals with all branches of pure and applied philosophy, from logic and philosophy of nature to ethics and philosophy of religion.

Ratio continues in a new form philosophical aims pursued with uncommon seriousness and courage in the *Abhandlungen der Fries'schen Schule.* The Abhandlungen were established in 1847 by Ernst Friedrich Apelt and re-established in 1904 by Leonard Nelson. Their publication was interrupted in 1937, but, in 1957, Julius Kraft, a pupil of Nelson, revived this work by founding *Ratio.*

Ratio is published twice a year: June and December

Subscriptions
Institutions: £20.75 (UK) £25.50 (overseas) $55.00 (USA) $66.00 (Canada)
Individuals: £14.00 (UK) £17.25 (overseas) $38.00 (USA) $45.50 (Canada)

Orders to: Journals Department, Basil Blackwell Publisher Ltd.,
108 Cowley Road, Oxford OX4 1JF, England.

The German edition of *Ratio* is published by FELIX MEINER VERLAG,
Richardstrasse 47, 2000 Hamburg 76, West Germany.

Published by BASIL BLACKWELL, Oxford

THEORIA

A Swedish journal of philosophy
Edited by Peter Gärdenfors, Bengt Hansson
and Krister Segerberg

Volume 47, part 2
SPECIAL ISSUE ON MORAL PHILOSOPHY
L. Jonathan Cohen Who is starving whom?
Raymond A. Belliotti Negative and positive duties
Rex Patrick Stevens The impact of theodicy on Kant's conception of moral practice

THEORIA, which invites contributions in all fields of analytic philosophy, usually publishes three issues a year. Subscriptions and requests for back numbers should be sent to *Editor*, THEORIA, *Filosofiska institutionen*, *Kungshuset i Lundagård*, *S-223 50 Lund*, *Sweden*. Subscription for one year is SwCr 60:– or U.S. $ 16.–, for three years SwCr 160:– or U.S. $ 40.–.

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

PUBLISHER. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. US Mailing Agent, Expediters of the Printed Word Ltd, 527 Madison Avenue, Suite 1217, New York, NY. 10022, USA. Second Class postage paid at New York.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £11.00 (inland), £13.50 (overseas), $33.00 (USA), $39.50 (Canada); for individuals £7.00 (inland), £8.60 (overseas), $21.00 (USA), $25.00 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year—in January, March, June and October. Orders should be sent to the publisher.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable.